

GenCore version 5.1.6
Copyright (C) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2004, 11:19:33 ; Search time 128 seconds
(without alignments)
6874.661 Million cell updates/sec

Title: US-10-030-389-12
Perfect score: 1238
Sequence: 1 acatcttggaatggacta.....ttttgagggtccagtta 1238

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents NA:*
- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
 - 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
 - 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
 - 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
 - 5: /cgn2_6/ptodata/1/ina/pctus_COMB.seq.*
 - 6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	226.2	18.3	1600	2	US-08-933-750C-50
2	226.2	18.3	1600	3	US-09-234-613-50
3	96.2	7.8	471	4	US-09-621-976-1305
4	82.2	5.0	708	4	US-09-270-767-13081
5	62	5.0	8607	4	US-10-204-708-72
6	60.8	4.9	396	4	US-09-640-173-53
7	60.8	4.9	396	4	US-09-713-550-53
8	60.8	4.9	396	4	US-09-825-294-53
9	60.8	4.9	396	4	US-09-970-966-53
10	58	4.7	2447	2	US-09-014-969-14
11	58	4.7	7218	1	US-08-232-463-14
12	57	4.6	240	1	US-08-628-417-6
13	57	4.6	582	4	US-09-787-292-3
14	56.4	4.6	441	4	US-09-601-537-10
15	56.4	4.6	2262	4	US-09-311-021-171
16	56.4	4.6	3275	4	US-09-370-838-151
17	56.4	4.6	3275	4	US-09-854-133-151
18	56.4	4.6	4121	4	US-09-601-537-9
19	56.2	4.5	1798	3	US-09-797-906-1
20	56	4.5	1024	4	US-09-328-475C-50
21	56	4.5	1985	4	US-09-907-794A-212
22	56	4.5	1985	4	US-09-905-125A-212
23	56	4.5	1985	4	US-09-902-775A-212
24	56	4.5	1985	4	US-09-906-700-212
25	56	4.5	1985	4	US-09-903-603A-212
26	55.8	4.5	5152	4	US-10-204-708-47
27	55.4	4.5	1020	4	US-09-328-475C-43

C 28	55.4	4.5	1813	3	US-09-071-224-3	Sequence 3, Appli
C 29	55.4	4.5	6671	1	US-08-280-443-1	Sequence 1, Appli
C 30	55.4	4.5	6671	1	US-08-457-459-1	Sequence 1, Appli
C 31	55.4	4.5	6671	1	US-08-555-678-1	Sequence 1, Appli
C 32	55.4	4.5	6671	5	PCT-US95-02275-1	Sequence 1, Appli
C 33	55.2	4.5	1798	2	US-08-557-128-12	Sequence 12, Appli
C 34	54.6	4.4	249	4	US-09-621-976-1322	Sequence 1322, Ap
C 35	54.6	4.4	396	4	US-09-640-173-33	Sequence 33, Appli
C 36	54.6	4.4	396	4	US-09-713-550-33	Sequence 33, Appli
C 37	54.6	4.4	396	4	US-09-825-294-33	Sequence 33, Appli
C 38	54.6	4.4	396	4	US-09-970-966-33	Sequence 3, Appli
C 39	54.6	4.4	2634	4	US-09-463-238-3	Sequence 1, Appli
C 40	54.4	4.4	1696	4	US-09-835-811-1	Sequence 93, Appli
C 41	54.4	4.4	10144	4	US-10-204-708-93	Sequence 3, Appli
C 42	54.2	4.4	2146	4	US-10-003-392-3	Sequence 3, Appli
C 43	54.2	4.4	2394	4	US-09-800-729-33	Sequence 1, Appli
C 44	54.2	4.4	2674	3	US-09-817-180-1	Sequence 1, Appli
C 45	54.2	4.4	2674	4	US-10-003-395-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-933-750C-50
; Sequence 50, Application US/08933750C
; Patent No. 5932442
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESSES:
; ADDRESSES: Insite Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,750C
; FILING DATE: September 23, 1997
; CLASSIFICATION: 536
; PRIOR APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1600 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: U937NOT01

US-09-270-767-13081
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13081
; LENGTH: 708
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
Query Match 5.0%; Score 62.2; DB 4; Length 708;
Best Local Similarity 57.4%; Pred. No. 9.2e-07;
Matches 112; Conservative 0; Mismatches 83; Indels 0; Gaps 0;
QY 963 GTTTTGTAAATTTTTCAGCAAAATTTTATACAGTGCATTAAATTCGGGAGAGATAC 1022
DB 657 GTTTTGTAAATTTTTCAGCAAAATTTTCCCGCCCGCCCAAAAATTTTGGGTTTCC 598
QY 1023 TGTCGAGAAATTAATGCATACATTTTGTACAAATTTCCCTTTTGGGGTACGTTTG 1082
DB 597 CCCCAAAAACATAATCCCGCCCGCCCGGGAATTTTATTTTATTTTATTTT 538
QY 1083 GTTTTGTAAATTTTTCAGCAAAATTTTATACAGTGCATTAAATTCGGGAGAGATAC 1142
DB 537 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 478
QY 1143 GTTGTATTTTGGCTT 1157
DB 477 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT

RESULT 6
US-09-640-173-53
; Sequence 53, Application US/09640173
; Patent No. 6613515
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: OVARIAN TUMOR SEQUENCES AND
; TITLE OF INVENTION: METHODS OF USE THEREFOR
; FILE REFERENCE: 210121.484C2
; CURRENT APPLICATION NUMBER: US/09/640.173
; CURRENT FILING DATE: 2000-08-15
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 53
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(396)
; OTHER INFORMATION: n = A,T,C or G
US-09-640-173-53

Query Match 4.9%; Score 60.8; DB 4; Length 396;
Best Local Similarity 51.2%; Pred. No. 1.6e-06;
Matches 131; Conservative 0; Mismatches 125; Indels 0; Gaps 0;
QY 959 TTCAGTTTGTAAATTTTTCAGCAAAATTTTATACAGTGCATTAAATTCGGGAGAG 1018
DB 14 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 73
QY 1019 ATACTGTCCAGAAATTAATGCATACATTTTGTACAAATTCCTTTTGGGGTACGT 1078
DB 74 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 133
QY 1079 TTTGGTTTATTTTGTGTTTATTTTCTTTCGATTTTCTTTCCTTTTCTTTTCTTTT 1138
DB 134 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 193
QY 1139 GGTGTGTTTATTTTGTCTTCAATATCTCTCGCGGTTCGGAATTCGTAACAGTAA 1198
DB 194 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 253
QY 1199 TTGAATGTTCGAAAAG 1214
DB 254 TCANAAAAAGAAAG 269

RESULT 7
US-09-713-550-53
; Sequence 53, Application US/09713550
; Patent No. 6617109
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF OVARIAN CANCER

US-10-204-708-72
; Sequence 72, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204.708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 72
; LENGTH: 8607
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-72

Query Match 5.0%; Score 62; DB 4; Length 8607;
Best Local Similarity 51.4%; Pred. No. 3.6e-06;
Matches 143; Conservative 0; Mismatches 135; Indels 0; Gaps 0;
QY 922 TTCAGATGGCTTGACATCAAGTGAATGTACAGGTTTCAGTTTCTTATATTTTCA 981
DB 6825 TTTTAAAGTGTGGGATATAGATGTGATTTATCCGTTTATGTTTAAATAATTTTAA 6884
QY 982 GCAATTTTATACAGTGTCAATTTAATTTGGGAGAGGATACCTCCAGAAAATTAATGCA 1041

RESULT 5
US-10-204-708-72
; Sequence 72, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204.708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 72
; LENGTH: 8607
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-72

This Page Blank (uspto)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2004, 11:19:33 ; Search time 682 Seconds
(without alignments)
9529,008 Million cell updates/sec

Title: US-10-030-389-12

Perfect score: 1238

Sequence: 1 acatccttggaatgagacta.....ttttgtggagggtccagtta 1238

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N Geneseq_23Sep04:*
1: Geneseqn1980s:*
2: Geneseqn1980s:*
3: Geneseqn2000s:*
4: Geneseqn2000s:*
5: Geneseqn2000s:*
6: Geneseqn2000s:*
7: Geneseqn2000s:*
8: Geneseqn2000s:*
9: Geneseqn2000s:*
10: Geneseqn2000s:*
11: Geneseqn2000s:*
12: Geneseqn2000s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1238	100.0	1238	4	AAF31245 Human imm
2	1174.2	94.8	2143	12	ADP07317 Human Wtl
3	1116.6	90.2	2772	10	Ades4051 Human pro
4	725.2	58.6	755	3	AAA44466 Human sec
5	296.6	24.0	440	8	ABX50288 Bovine ES
6	249	20.1	322	10	ACD94748 Human col
7	231.4	18.7	2016	5	AA84773 DNA encod
8	226.2	18.3	1600	8	ABX63550 Human cdn
9	226.2	18.3	1622	6	ABL65182 Lung can
10	226.2	18.3	1691	8	ABX63552 Human cdn
11	226.2	18.3	1996	8	ABX63551 Human cdn
12	201.2	16.3	475	6	ABL79439 Human ova
13	184	14.9	1565	12	ADOS8729 DNA encod
14	167.8	13.6	2553	4	ABL10891 Drosophil
15	163.6	13.2	330	4	AA184750 Human pol
16	152.2	12.3	3028	8	AA84774 DNA encod
17	143.4	11.6	3900	8	ABX6451 Human cdn
18	143.4	11.6	3902	10	ADJ56454 Human cdn
19	125.4	10.1	313	12	ADH00046 Kidney cl
20	120.8	9.8	2642	11	ADM03477 Human cdn
21	120.8	9.8	81826	10	ADL13767 Osteoearth

22	120.8	9.8	81826	10	ADL13783	Adl13783 Osteoearth
23	120.8	9.8	201766	10	ADL13771	Adl13771 Osteoearth
24	115.2	9.3	501	6	AA561638	AA561638 Lung smal
25	108	8.7	481	9	ACH42385	ACH42385 Human foe
26	101.4	8.2	6916	4	ABL10890	ABL10890 Drosophil
27	100.2	8.1	479	5	AA574501	AA574501 DNA encod
28	93	7.5	558	6	AAQ97529	AAQ97529 Mouse ES
29	79.2	6.4	1136	3	AA468838	AA468838 Arabidops
30	75.4	6.1	34548	6	ABL70603	ABL70603 Chemical
31	75	6.1	501	6	AA561627	AA561627 Lung smal
32	75	6.1	33053	6	ABQ67005	ABQ67005 Human ang
33	72.4	5.8	6794	6	ABK31264	ABK31264 Signal tr
34	72.4	5.8	6794	6	ABL70219	ABL70219 Chemical
35	72.4	5.8	6794	6	AA561174	AA561174 Human gen
36	71.4	5.8	4324	6	AAQ67149	AAQ67149 Human ang
37	69	5.6	5823	6	ABK28382	ABK28382 DNA trans
38	68	5.5	12007	6	ABL32717	ABL32717 Human imm
39	67	5.4	18624	6	ABL33703	ABL33703 Human imm
40	66.8	5.4	411	4	AA187442	AA187442 Human pol
41	66.8	5.4	5413	4	AA546694	AA546694 Tumour su
42	66.6	5.4	411	8	ABX49356	ABX49356 Bovine ES
43	66.6	5.4	6668	6	ABL33696	ABL33696 Human imm
44	66	5.3	2381	6	ABL34337	ABL34337 Human imm
45	65.8	5.3	6306	4	AA545516	AA545516 Chemical

ALIGNMENTS

RESULT 1

AAF31245

ID AAF31245 standard; DNA; 1238 BP.

XX

AC AAF31245;

XX

DT 09-APR-2001 (first entry)

XX

DE Human immune response molecule (IMUN) coding sequence SEQ ID NO: 12.

XX

KW Human; IMUN; immune response molecule; autoimmune disorder;

KW inflammatory disorder; cell proliferation disorder; cancer; ss.

OS Homo sapiens.

XX

PN WO200102569-A2.

XX

PD 11-JAN-2001.

XX

PF 06-JUL-2000; 2000WO-US018505.

XX

PR 06-JUL-1999; 99US-0142572P.

PR

09-SEP-1999; 99US-0153170P.

XX

(INCY-) INCYTE GENOMICS INC.

XX

PA Tang YT, Yue H, Yang J, Azimzai Y, Baughn MR, Lu DAM;

XX

PI WPI; 2001-123113/13.

DR

P-PSDB; AAB67048.

XX

PT Novel human immune response molecules (IMUN), useful for diagnosing, treating and preventing disorders associated with abnormal expression of IMUN, e.g. Addison's disease, allergies, anemia, asthma and atherosclerosis.

XX

PS Claim 5; Page 90; 95pp; English.

XX

CC The present invention provides the protein and coding sequences for ten human immune response molecules (designated IMUN-1 to IMUN-10). The sequences can be used in the diagnosis, treatment and prevention of CC autoimmune and inflammatory disorders such as AIDS, atherosclerosis, CC asthma, allergies, Crohn's disease, multiple sclerosis, irritable bowel CC syndrome, psoriasis, rheumatoid arthritis and infections, and cell

CC proliferation disorders including arteriosclerosis, cirrhosis and cancer
 XX SQ Sequence 1238 BP; 363 A; 249 C; 302 G; 324 T; 0 U; 0 Other;

Query Match 100.0%; Score 1238; DB 4; Length 1238;

Best Local Similarity 100.0%; Pred. No. 2e-277;

Matches 1238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACATCCTTGGATGCGACTAGCAACCAAGGACAGATATTCAGAGTGCTACTCAAA 60
 DB 1 ACATCCTTGGATGCGACTAGCAACCAAGGACAGATATTCAGAGTGCTACTCAAA 60
 QY 61 TCCAGTACTCAAGCAAGTCCAGCAGCTAGCGTTGCCAATTTAGATCAACAATGGTAG 120
 DB 61 TCCAGTACTCAAGCAAGTCCAGCAGCTAGCGTTGCCAATTTAGATCAACAATGGTAG 120
 QY 121 ACCGAGCGATCAACTGTTTTTCTTAAATGAAAGTGAACTGGAACAGACTAAAGACA 180
 DB 121 ACCGAGCGATCAACTGTTTTTCTTAAATGAAAGTGAACTGGAACAGACTAAAGACA 180
 QY 181 AACTGGAAACAAGCCCAAAATGAATGAGTGCCTGGAAAGTTTACGCCCTGATAGCCAAACAG 240
 DB 181 AACTGGAAACAAGCCCAAAATGAATGAGTGCCTGGAAAGTTTACGCCCTGATAGCCAAACAG 240
 QY 241 GGAAAAGTTAATGGCGAAGTGTGAATGCTTATCCAGGAGAAATCAAGAGCTTGGAAAGC 300
 DB 241 GGAAAAGTTAATGGCGAAGTGTGAATGCTTATCCAGGAGAAATCAAGAGCTTGGAAAGC 300
 QY 301 AGCTGTCCAGGGAGCTATTTCACAACTTGAAGCAGAGTTGGCTTTACAGAAGAAATACA 360
 DB 301 AGCTGTCCAGGGAGCTATTTCACAACTTGAAGCAGAGTTGGCTTTACAGAAGAAATACA 360
 QY 361 GTGAGGAGCTTAAAGCAGTCCAGATGAATGATCAATCATCCAGCTTGTATGAAG 420
 DB 361 GTGAGGAGCTTAAAGCAGTCCAGATGAATGATCAATCATCCAGCTTGTATGAAG 420
 QY 421 AAGTAGAGGGTATCCAGAGTACCATCTTGTCTGACAGCAGCTGTAAGGAGACACGCC 480
 DB 421 AAGTAGAGGGTATCCAGAGTACCATCTTGTCTGACAGCAGCTGTAAGGAGACACGCC 480
 QY 481 AGCAGTTGGCTCAGTACAGCAGCAGAGTCTCAGGCTCTGCCCCAGAGTACCAGCAGGA 540
 DB 481 AGCAGTTGGCTCAGTACAGCAGCAGAGTCTCAGGCTCTGCCCCAGAGTACCAGCAGGA 540
 QY 541 CTACAGCTTCTGAACCTGTAGAACAGTCCAGAGCCCAAGTAAGAGTGCAGTCGCTGA 600
 DB 541 CTACAGCTTCTGAACCTGTAGAACAGTCCAGAGCCCAAGTAAGAGTGCAGTCGCTGA 600
 QY 601 CAAACGGACCAAGTAATGGTAGCTCTCCCGCCAGAGAGCTCTGGGTCTGGATTTTACA 660
 DB 601 CAAACGGACCAAGTAATGGTAGCTCTCCCGCCAGAGAGCTCTGGGTCTGGATTTTACA 660
 QY 661 GGGAGGGCAACAACCGAAGATGACTTCTCTTCCAGGGAAATGTAATGAAGTCT 720
 DB 661 GGGAGGGCAACAACCGAAGATGACTTCTCTTCCAGGGAAATGTAATGAAGTCT 720
 QY 721 CCAACAGCTCAGAGCAGAGAACTGGCAGAGAGTAGTGTACGTAAATCAACTCAGTG 780
 DB 721 CCAACAGCTCAGAGCAGAGAACTGGCAGAGAGTAGTGTACGTAAATCAACTCAGTG 780
 QY 781 CGGGGTATGAAGTGTAGATCTTCCACGGGAGTGAATACTCTCTCACACCACTCAAA 840
 DB 781 CGGGGTATGAAGTGTAGATCTTCCACGGGAGTGAATACTCTCTCACACCACTCAAA 840
 QY 841 ATGACACAGCTCCAGTCAATGACCTCAAGAGAGAAAGCAGTGAAGTGAAGTAATC 900
 DB 841 ATGACACAGCTCCAGTCAATGACCTCAAGAGAGAAAGCAGTGAAGTGAAGTAATC 900
 QY 901 GAAGTGTGGGTCCCGCCAGTTCAGAAATGGCTTGGACTCAAGTGTAAATGTACAGGGTT 960
 DB 901 GAAGTGTGGGTCCCGCCAGTTCAGAAATGGCTTGGACTCAAGTGTAAATGTACAGGGTT 960
 QY 961 CAGTTTGTAAATATTTTTCAGCAAAATTTTATACAGTGTCTATTTAATTTGGAGAGGAT 1020

DB 961 CAGTTTGTAAATATTTTTCAGCAAAATTTTATACAGTGTCTATTTAATTTGGAGAGGAT 1020
 QY 1021 ACTGTCGAGAAATTAATGATACATCTTTGTGTCACAAATTTGCCCTTTTGTGGGTGACGTTT 1080
 DB 1021 ACTGTCGAGAAATTAATGATACATCTTTGTGTCACAAATTTGCCCTTTTGTGGGTGACGTTT 1080
 QY 1081 TGGTTTTTTTTTGTGTTTTTTTTTCTTTGGATTTTTTTTTTCTTTTAAAGAGGTGG 1140
 DB 1081 TGGTTTTTTTTTGTGTTTTTTTTTCTTTGGATTTTTTTTTTCTTTTAAAGAGGTGG 1140
 QY 1141 TTGTTGTTTTTTTGTCTTCAATCTTCTCGCGCTTGGGAATTTAAGATTAATTAATCTTT 1200
 DB 1141 TTGTTGTTTTTTTGTCTTCAATCTTCTCGCGCTTGGGAATTTAAGATTAATTAATCTTT 1200
 QY 1201 GAATGTTCCAAAGGGCATTTTGTGGGGTCCAGTTA 1238
 DB 1201 GAATGTTCCAAAGGGCATTTTGTGGGGTCCAGTTA 1238

RESULT 2

ADP07317

ID ADP07317 standard; DNA; 2143 BP.

XX AC ADP07317;

XX DT 29-JUL-2004 (first entry)

XX DE Human WT1 DNA.

XX KW ds; proliferative disease; breast; methylation; CpG; bisulfite; human.

XX OS Homo sapiens.

XX PN DE10255104-A1.

XX PD 11-MAR-2004.

XX PF 26-NOV-2002; 2002DE-01055104.

XX PR 27-AUG-2002; 2002DE-01039313.

XX PA (EPIC-) EPIGENOMICS AG.

XX PI Maier S;

XX WPI; 2004-284340/27.
 GENBANK; NM_004906.

Analyzing proliferative diseases of breast cells, useful e.g. for
 diagnosis, prognosis and treatment of breast cancer, by determining
 methylation status of specific genes.

Claim 1; Page: 22pp; German.

This invention describes a novel method of analysing proliferative
 diseases of breast cells by determining the methylation status of certain
 genes. The invention also describes nucleic acids or their complements,
 oligomers, especially oligonucleotides or peptide nucleic acid oligomers,
 that hybridise to, or are identical with, any of the nucleic acids, the
 preparation of an oligomer array for analysing proliferative diseases of
 breast cells that are associated with the methylation status of CpG
 dinucleotides of the genes by bonding at least one oligomer to a solid
 phase and a kit comprising a bisulfite reagent and the oligomer. The
 nucleic acids are genomic sequences (5'- and/or regulatory and/or CpG-
 rich regions). The base sequence of the oligomer includes at least one
 CpG island, especially with C in the middle third of the sequence. The
 process involves a genomic DNA sample treated chemically, specifically
 with a bisulfite reagent, to convert non-methylated C to uracil or some
 other base having base-pairing properties different from those of C.
 Fragments of the treated DNA are then amplified, using primers and a
 polymerase and the methylation status of the genomic CpG dinucleotides is
 determined by analysis of the amplicons, particularly by hybridisation to

the oligomer, optionally with extension of the hybridised oligomer by at least one base, or detection is by sequencing. The amplification may use methylation-specific primers. Alternatively, genomic DNA is extracted from a sample and digested with methylation-specific restriction enzymes, then the digestion fragments detected, optionally after amplification. In either method, more than 10 fragments of 100-200 bp are amplified in a single reaction vessel, using a heat-resistant DNA polymerase in PCR. The amplicons carry detectable markers, e.g. fluorophores, radioisotopes and/or releasable fragments of known mass that can be detected by mass spectrometry. The method is used for characterisation, classification, differentiation, staging, phase-estimation, diagnosis and/or therapy of proliferative diseases of breast cells. The method provides very specific classification of proliferative diseases, allowing better treatment. It can both characterise methylation status and detect single-nucleotide polymorphisms. This sequence represents human gene used to illustrate the method of the invention. NOTE: This sequence does not appear in the printed specification but has been retrieved from Genbank.

Sequence 2143 BP; 619 A; 431 C; 498 G; 595 T; 0 U; 0 Other;

Query Match 94.8%; Score 1174.2; DB 12; Length 2143;
Best Local Similarity 97.8%; Pred. No. 1.6e-262;
Matches 1212; Conservative 0; Mismatches 23; Indels 4; Gaps 2;
QY 1 ACATCCCTGGAATCGACTAGCAACCAAGGAGGACAGATATTCAGAGTGTACTACTCAA 60
DB 443 ACATCCCTTGTATCGGACTAGCAACCAAGGAGGAGTGTACTACTCAA 502
QY 61 TCCAGTACCTCAAGCAAGTCAGAGCCTAGCGTTGCCCACTTAGATCAACAATGGTAG 120
DB 503 TCCAGTACCTCAAGCAAGTCAGAGCCTAGCGTTGCCCACTTAGATCAACAATGGTAG 562
QY 121 ACCAGCGATCAACTGTTTTCCTAAAATGAAGGTGAAGTGAAGTGAAGTGAAGTGAAG 180
DB 563 ACCAGCGATCAACTGTTTTCCTAAAATGAAGGTGAAGTGAAGTGAAGTGAAGTGAAG 622
QY 181 AACTGGAAACAGCCCAAAATGAATGAGTGTGCTGGAAAGTTTACGCCGTGATAGCCAAACAG 240
DB 623 AACTGGAAACAGCCCAAAATGAATGAGTGTGCTGGAAAGTTTACGCCGTGATAGCCAAACAG 682
QY 241 GGAAAAAGTTAATGGCGAAGTGTGATGCTTATCCAGGAGAAATCAAGAGCTTGGAGGC 300
DB 683 GGAAAAAGTTAATGGCGAAGTGTGATGCTTATCCAGGAGAAATCAAGAGCTTGGAGGC 742
QY 301 AGCTGTCCAGCGGAGTATTGCACAACTTGAAGCAGAGTGTGCTTTACAGAAAGAAATACA 360
DB 743 AGCTGTCCAGCGGAGTATTGCACAACTTGAAGCAGAGTGTGCTTTACAGAAAGAAATACA 802
QY 361 GTGAGGAGCTTAAAAGCAGTCAGATGAACCTGAATGATCTCATCATCAGCTTGATGAAG 420
DB 803 GTGAGGAGCTTAAAAGCAGTCAGATGAACCTGAATGATCTCATCATCAGCTTGATGAAG 862
QY 421 AAGTAGAGGTTATGAGAGTACCACTTCTAGTCTGACAGCAGCTGAAGAGAGACAGCC 480
DB 863 AAGTAGAGGTTATGAGAGTACCACTTCTAGTCTGACAGCAGCTGAAGAGAGACAGCC 922
QY 481 AGCAGTTGGCTCAGTACCAGCAGCAGCAGTCTCAGGCTCTGCCCAAGTACCAGCAGGA 540
DB 923 AGCAGTTGGCTCAGTACCAGCAGCAGCAGTCTCAGGCTCTGCCCAAGTACCAGCAGGA 982
QY 541 CTACAGCTTCTGAACCTGTAGAACAGTCAGAGGCCACAAAGTAAAGACTGCGTCTGA 600
DB 983 CTACAGCTTCTGAACCTGTAGAACAGTCAGAGGCCACAAAGTAAAGACTGCGTCTGA 1042
QY 601 CAAACGGACCAAGTAAATGTTAGTCTCTCCGCGCAGGAGCAGCTCTGGGTCTGGATTTCACA 660
DB 1043 CAAACGGACCAAGTAAATGTTAGTCTCTCCGCGCAGGAGCAGCTCTGGGTCTGGATTTCACA 1102
QY 661 GGGAGGGCAACACACCGAAGATGACTTTCTCTTCTCCAGGGAATGCTTAATAGTCTCT 720
DB 1103 GGGAGGGCAACACACCGAAGATGACTTTCTCTTCTCCAGGGAATGCTTAATAGTCTCT 1162
QY 721 CCAACAGCTCAGAGGAGAGAACTGGCAGGAGGAGTGTAGTGTACGTAATCAATCACTAGT 780

DB 1163 CCAACAGCTCAGAGGAGAGAACTGGCAGAGAGGTAGTGTAGTAAATCAATCACTAGTG 1222
QY 781 CGGGGTATGAAAGTGTAGACTCTCCACGGGCGAGTGAATACTCTCTCACACCAATCAA 840
DB 1223 CGGGGTATGAAAGTGTAGACTCTCCACGGGCGAGTGAATACTCTCTCACACCAATCAA 1282
QY 841 ATGACACAGACTCCAGTCTCAGTCTCAGCCCTCAAGAGGAGAGCAGTGTAGTGGGAAAGTAAATC 900
DB 1283 ATGACACAGACTCCAGTCTCAGTCTCAGCCCTCAAGAGGAGAGAGCAGTGTAGTGGGAAAGTAAATC 1342
QY 901 GAACTGTGGGTTCGCGCCACAGCTTCAGAAATGCTTGGACTCAAGTGTAAATGTACAGGGTT 960
DB 1343 GAACTGTGGGTTCGCGCCACAGCTTCAGAAATGCTTGGACTCAAGTGTAAATGTACAGGGTT 1402
QY 961 CAGTTTGTAAATATTTTTTCAGCAATTTTTTATACAGTGTCTATTAATTTGGGAGAGGAT 1020
DB 1403 CAGTTTGTAAATATTTTTTCAGCAATTTTTTATACAGTGTCTATTAATTTGGGAGAGGAT 1462
QY 1021 ACTGTCCAGAAATTAATGCATATCTTTGTCAATTTGCTTTTGTGGGTGTACGTTT 1080
DB 1463 ACTGTCCAGAAATTAATGCATATCTTTGTCAATTTGCTTTTGTGGGTGTACGTTT 1522
QY 1081 TCGTTTTTTTGTGTTTTTTTCTTGGATTTTTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTT 1579
DB 1523 TCGTTTTTTTGTGTTTTTTTCTTGGATTTTTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTT 1579
QY 1141 TTGTTGTTTTTTTGTCTCAATACTTCTGCGCTTGGAAATTTGAACAGTTAAATTAATTAATTT 1200
DB 1580 TTTTTTTTTTTTTGCTTCAATACTTCTGCGCTTGGAAATTTGAACAGTTAAATTAATTAATTT 1639
QY 1201 GAATGTGTC-AAAAGGCGATTTTGTGGAGGTCCTAGTTA 1238
DB 1640 GAATGTGTC-AAAAGGCGATTTTGTGGAGGTCCTAGTTA 1678

RESULT 3

AD54051
ID ADE54051 standard; cDNA; 2772 BP.

XX AC ADE54051;

XX DT 29-JAN-2004 (first entry)

XX DE Human prostate cancer cDNA #398.

XX DE Human; prostate cancer; ss; cDNA combination; differential expression;
KW gene.

XX OS Homo sapiens.

XX PN US2003190640-A1.

XX PD 09-OCT-2003.

XX PF 29-MAY-2002; 2002US-00252157.

XX PR 31-MAY-2001; 2001US-0295048P.

XX PA (FARI/) FARIS M.

XX PA (PEAR/) PEARSON C I.

XX PI Faris M, Pearson CI;

XX DR WPI; 2003-831619/77.

XX PT New combination comprising cDNAs that are differentially expressed in
PT prostate cancer, useful for diagnosing, treating or monitoring the
PT progression of treatment of prostate cancer.

XX PS Claim 1; SEQ ID NO 398; 42pp; English.

XX CC The invention relates to a combination comprising a number of cDNAs

expressed in prostate cancer. The invention also relates to a method for detecting differential expression of one or more cDNAs in a sample containing nucleic acids by hybridising a substrate with the nucleic acids, thus forming one or more hybridisation complexes, detecting hybridisation complex formation and comparing the complexes formed with standard complexes, where differences between the standard and the sample complex formation indicate differences in the expression of cDNAs in the sample. The differential expression is diagnostic of prostate cancer. The invention also relates to proteins and antibodies related to the cDNAs. The combination is useful for diagnosing, treating or monitoring the progression of treatment of prostate cancer. The antibodies are useful for detecting prostate cancer. This sequence represents a human prostate cancer cDNA of the invention.

XX SQ Sequence 2772 BP; 740 A; 680 C; 654 G; 688 T; 0 U; 0 Other;

Query Match 90.2%; Score 1116.6; DB 10; Length 2772;
Best Local Similarity 99.2%; Pred No. 4.1e-249;
Matches 1122; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 ACATCCTTGGAAATGCGACTAGCAACCAAGGACAGATATTCAAGAGTGTACTACTCAAA 60
DB 1641 ACATCCTTGGAAATGCGACTAGCAACCAAGGACAGATATTCAAGAGTGTACTACTCAAA 1700
QY 61 TCCAGTACTCAGCAAGTCCAGCAGCTAGCGTTGCCCACTTAGATCAACAATGGTAG 120
DB 1701 TCCAGTACTCAGCAAGTCCAGCAGCTAGCGTTGCCCACTTAGATCAACAATGGTAG 1760
QY 121 ACCAGCGATCAACTGTTTTCTTAAATGAAGGTGAAGTGAACAGACTAAAGACA 180
DB 1761 ACCAGCGATCAACTGTTTTCTTAAATGAAGGTGAAGTGAACAGACTAAAGACA 1820
QY 181 AACTGGAAACAGCCAAATGAATGAATGAGTGGTGAAGTTTACGCTGATAGCAACAG 240
DB 1821 AACTGGAAACAGCCAAATGAATGAGTGGTGAAGTTTACGCTGATAGCAACAG 1880
QY 241 GGAAAGAGTTAATGGCAAGTGTGCAATGCTTATCCAGGAGATCAAGAGCTTGGAGGC 300
DB 1881 GGAAAGAGTTAATGGCAAGTGTGCAATGCTTATCCAGGAGATCAAGAGCTTGGAGGC 1940
QY 301 AGCTGTCCAGGACGTAATTGCAACAATGAAGCAGAGTTGGCTTTACAGAAGAAATACA 360
DB 1941 AGCTGTCCAGGACGTAATTGCAACAATGAAGCAGAGTTGGCTTTACAGAAGAAATACA 2000
QY 361 GTGAGGAGCTTAAAGCAGTCAAGATCAATGATGATCTCATCCTCAGCTTGTGTAAG 420
DB 2001 GTGAGGAGCTTAAAGCAGTCAAGATCAATGATGATCTCATCCTCAGCTTGTGTAAG 2060
QY 421 AAGTAGAGGATTCAGAGTACCAATCTAGTTCTGACAGCAGCTGAAGGAGACACGCC 480
DB 2061 AAGTAGAGGATTCAGAGTACCAATCTAGTTCTGACAGCAGCTGAAGGAGACACGCC 2120
QY 481 AGCAGTTGGCTCAGTACCAAGCAGCAGTCTCAGGCTCTGCCCAAGTACCAGCAGGA 540
DB 2121 AGCAGTTGGCTCAGTACCAAGCAGCAGTCTCAGGCTCTGCCCAAGTACCAGCAGGA 2180
QY 541 CTACAGCTTCTGACCTGTAGAAACAGTCAGAGCCCAAGTAAGACTGCAGTCGCTCA 600
DB 2181 CTACAGCTTCTGACCTGTAGAAACAGTCAGAGCCCAAGTAAGACTGCAGTCGCTCA 2240
QY 601 CAACAGCACAAGTAATGGTAGCTCTCCGCCAGAGGAGCTGGGTCTGGAATTTACA 660
DB 2241 CAACAGCACAAGTAATGGTAGCTCTCCGCCAGAGGAGCTGGGTCTGGAATTTACA 2300
QY 661 GGGAGGGCAACACAAACGAAGATGACTTCTCTTCTCCAGGGAATGTAATAGTCT 720
DB 2301 GGGAGGGCAACACAAACGAAGATGACTTCTCTTCTCCAGGGAATGTAATAGTCT 2360
QY 721 CCACAGCTCAGAGGAGAACTGGCAGAGAGGTAGTGTACGTAATCAACTCAGTG 780
DB 2361 CCACAGCTCAGAGGAGAACTGGCAGAGAGGTAGTGTACGTAATCAACTCAGTG 2420
QY 781 CGGGGTATGAAGTGTAGACTCTCCCAAGGAGTGAAGTCTCTCACACCAACTCA 840

DB 2421 CGGGGTATGAAGTGTAGACTCTCCACGGGAGTGAAGTCTCTCACACCAACTCAA 2480
QY 841 ATGACACAGACTCCAGTCTAGTATGACCCCTCAAGAGGAGAAAGCAGTGTAGTGGAAAGTAAATC 900
DB 2481 ATGACACAGACTCCAGTCTAGTATGACCCCTCAAGAGGAGAAAGCAGTGTAGTGGAAAGTAAATC 2540
QY 901 GAACTGTGGTTCCTCCGACAGTTCAGATGGCTTGGACTCAAGTGTAAATGTACAGGGTT 960
DB 2541 GAACTGTGGTTCCTCCGACAGTTCAGATGGCTTGGACTCAAGTGTAAATGTACAGGGTT 2600
QY 961 CAGTTTGTAAATATTTTTCAGCAAAATTTTATACAGTGTCAATTTAATTTGGGAGAGAT 1020
DB 2601 CAGTTTGTAAATATTTTTCAGCAAAATTTTATACAGTGTCAATTTAATTTGGGAGAGAT 2660
QY 1021 ACTGTCCGAAATTAATGCATACCTTTGTCAAAATTTGCCCTTTTGTGGGTGACGTTT 1080
DB 2661 ACTGTCCGAAATTAATGCATACCTTTGTCAAAATTTGCCCTTTTGTGGGTGACGTTT 2720
QY 1081 TGGTTTTTTTGTGTTTTTTTCTTGGATTTTTTTTCTTTTCTTTT 1131
DB 2721 TGGTTTTTTTGTGTTTTTTTCTTGGATTTTTTTTCTTTTCTTTT 2771
RESULT 4
AAA44466
ID AAA44466 standard; cDNA; 755 BP.
XX AC AAA44466;
XX DT 21-AUG-2000 (first entry)
XX DE Human secreted expressed sequence tag SEQ ID NO:1041.
XX KW Human; mouse; chicken; rat; secreted expressed sequence tag; sEST;
XX KW expressed sequence tag; EST; probe; chemotactic; proliferative;
XX KW immunomodulatory; hematopoietic; chemokinetic; analgesic; haemostatic;
XX KW thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
XX KW antiviral; antidiabetic; antisthmatic; vulnary; antiparkinsonian;
XX KW antitumor; osteopathic; neuroprotective; nootropic; antipsoriatic;
XX KW cerebroprotective; anticonvulsant; antidepressant; gene therapy; vaccine;
XX KW autoimmune disorder; multiple sclerosis; allergic condition;
XX KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
XX KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
XX KW central nervous system disorder; Alzheimer's disease; stroke;
XX KW Parkinson's disease; Huntington's disease; coagulation disorder;
XX KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease; tumour;
XX KW infection; depression; psoriasis; ss.
OS Homo sapiens.
XX WO2000021991-A1.
XX PN 20-APR-2000.
XX PD 15-OCT-1999; 99WO-US024206.
XX PF 15-OCT-1998; 98US-010436P.
XX PR (GENY) GENETICS INST INC.
XX PA Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
XX PI Merberg D, Treacy M, Bowman MR;
XX XX WPI; 2000-317938/27.
XX DR Isolated polynucleotides, and encoded proteins, comprising secreted
XX PT expressed sequence tags (sESTs), useful for treating various disorders
XX PT such as autoimmune, infectious, and central nervous system disorders.
XX PS Claim 1; Page 470-471; 803pp; English.
XX CC AAA44466 to AAA45925 represent specifically claimed secreted expressed


```
CC seqdata.uspto.gov/sequence.html?DocID=20020137160
XX Sequence 440 BP; 123 A; 85 C; 92 G; 140 T; 0 U; 0 Other;
XX
Query Match      24.0%; Score 296.6; DB 8; Length 440;
Best Local Similarity 89.5%; Pred. No. 6.2e-59;
Matches 332; Conservative 0; Mismatches 34; Indels 5; Gaps 1;
QY 742 CTGGCAGAGGAGTGTAGTTCGTTACGTPAAATCAACTCAGTCGCGGGTATGAAAGTGTAGACT 801
Db 1 CTGGCAGAGGAGTGTAGTTCGTTACGTPAAATCAACTCAGTCGCGGGTATGAAAGTGTAGACT 60
QY 802 CTCCACCGGCGAGTCAAAACTCTCTCACACACCACCAATCAATACACAGACTCCAGTCATG 861
Db 61 CTCCACCGGCGAGTCAAAACTCTCTCACACACCACCAATCAATACACAGACTCCAGTCATG 120
QY 862 ACCCTCAAGAGAGAGAAAGCAGTGTGGGAAAGGTAAATCGAACTGTGGTTCCTCCGCCACG 921
Db 121 ACCGCAAGAGAGAGAAACCTGTGAGTGGAAAGGTAAACCGAACTGGCGCTCCGCCACG 180
QY 922 TTCAGATGGCTTGGACTCAAGTGTAAATGTACAGGGTTCAGTTTTGT-----AATATT 976
Db 181 TTCAGATGGCTTGGACTCAAGTGTAAATGTACAGGGTTCAGTTTTGTAAAAAATAAAT 240
QY 977 TTTCAGCAAAATTTTATACAGTGTCAATTAATTTGGGAGAGGATACGTCCAGAAATTA 1036
Db 241 TTTCAGCAAAATTTTATACAGTGTCAATTAATTTGGGAGAGGATACGTCCAGTAAATTA 300
QY 1037 ATGCATCTTTTGTACAAATTTGCTTTTGTGGGTGTACGTTTGTGTTTTTTTGTG 1096
Db 301 ATGCATCTTTTGTACAAATTTGCTTTTGTGGGTGTACGTTTGTGTTTTTTTGTG 360
QY 1097 TTTTCTTTCTT 1107
Db 361 TTTCTTTGCTT 371
RESULT 6
ID ACD94748 standard; cdna; 322 BP.
XX ACD94748;
XX
XX 23-SEP-2003 (first entry)
XX Human colon cancer cell expressed cdna #3160.
XX
XX Open reading frame detection; genome sequencing; colon cancer;
KW breast cancer; population genome analysis; genetic shift; cancer;
KW antibiotic resistance; antibiotic non-tolerance; congenital disease;
KW agriculture; food crop genome; resistance gene; retrovirus;
KW influenza virus; eukaryotic pathogen detection; trypanosome; Plasmodium;
KW gene; ss.
XX
XX Homo sapiens.
XX
XX US2002155438-A1.
XX
XX 24-OCT-2002.
XX
XX 27-SEP-1999; 99US-00406117.
XX
XX 20-NOV-1998; 98US-00196716.
XX
XX (SIMP/) SIMPSON A J G.
XX (NETO/) NETO E D.
XX (BREN/) BRENTANI R R.
XX
XX Simpson AJG, Neto ED, Brentani RR;
XX
XX WPI; 2003-182626/19.
XX
XX Determining open reading frames of genome of an organism e.g. a human
PT suffering from cancer involves use of single oligonucleotide primer at
PT low stringency for preparing single-stranded cDNA from mRNA of
PT individual.
XX
XX Example 9; Page 465; 959pp; English.
XX
XX The invention describes a method of determining open reading frames in
XX the genome of organism, comprising contacting mRNA from cell of organism
XX with a single oligonucleotide primer (I) at low stringency, preparing
XX single-stranded cDNA by reverse transcribing mRNA with (I), amplifying
XX cDNA, sequencing the product, and repeating the contacting, preparing
XX and amplifying steps with different primers and sequencing, resulting
XX nucleic acids. The method is useful for: determining that a known
XX nucleotide sequence from a genome of an organism corresponds to a
XX nucleic acid molecule from a genome of an organism; and for sequencing
XX all or part of a genome of an organism. mRNA is obtained from mammalian
XX or human cell which is associated with a pathological condition e.g. a
XX colon cancer or breast cancer cell. The method is useful for analyses of
XX populations of subjects and can be used to carry out genetic analyses of
XX large or small populations. further, it can be used to study living
XX systems to determine if, e.g. there have been genetic shifts which render
XX an individual or population more or less likely to be afflicted with
XX diseases such as cancer, to determine antibiotic resistance or non-
XX tolerance, and so forth. The method can also be used in the study of
XX congenital diseases, and the risk of affliction to a foetus, as well as
XX the study of whether the conditions are likely to be passed to offspring
XX through ova or sperm. The analyses for pathological conditions can be
XX carried out in all animals, plants, birds, fish, etc. Using this method,
XX in the area of agriculture, for example the genomes of food crops can be
XX studied to determine if resistance genes are present, defects in plant
XX genomes can also be studied in this way. Similarly, the method permits
XX determination of the pathogens which integrate into the genome, such as
XX retroviruses and other integrating viruses such as influenza virus, have
XX undergone shifts or mutations, which may require different approaches to
XX therapy. This method is also applied to eukaryotic pathogens, such as
XX trypanosomes, different types of Plasmodium, etc. The method essentially
XX eliminates sequencing of non-coding portions. This sequence represents a
XX polynucleotide isolated from human colon cancer cell cdna library
SQ Sequence 322 BP; 63 A; 76 C; 73 G; 106 T; 0 U; 4 Other;
Query Match      20.1%; Score 249; DB 10; Length 322;
Best Local Similarity 94.7%; Pred. No. 6.3e-48;
Matches 266; Conservative 0; Mismatches 14; Indels 1; Gaps 1;
QY 2 CATCCTTGGATGGACTAGCAACCAAGGACACAGATATTCAGAGTGTACTCTCAAT 61
Db 280 CATCCTTGGATGGACTAGCAATNCAA-GGTCAAGAGATGCAAGAGTGTACTCTCAAT 222
QY 62 CCAGTACCTCAAGCAAGTCCAGCAGCCTAGCGTTGCCCACTTAGATCAACAATGGTAGA 121
Db 221 CCAGTACCTCAAGCAAGTCCAGCAGCAGCGGTTGNCATCTGAGATCAACAATGGTAGT 162
QY 122 CCCAGCGATCAACTTTGTTTCTTAAATAATGAAAGGTGAACTGGAACAGACTAAAGACAA 181
Db 161 CCCAGCGATCAACTTTGTTTCTTAAATAATGAAAGGTGAACTGGAACAGACTAAAGACAA 102
QY 182 ACTGGACAGAGCCCAAAATGAACTGAGTGCCTGGAGTTTACGCCCTATAGCCCAACAGG 241
Db 101 ACTGGACAGAGCCCAAAATGAACTGAGTGCCTGGAGTTTACGCCCTATAGCCCAACAGG 42
QY 242 GAAAAAGTTAATGGCGAAGTGTGCAATGCTTATCCAGGAGA 282
Db 41 GAAAAAGTTAATGGCGAAGTGTGCAATGCTTATCCAGGAGA 1
RESULT 7
AAS84773
ID AAS84773 standard; cdna; 2016 BP.
XX AAS84773;
XX
```

DT XX 13-FEB-2002 (first entry)
DE XX DNA encoding novel human diagnostic protein #20577.
KW Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder; ss.
OS Homo sapiens.
PN W0200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US0008631.
XX 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
PA Dmanac RT, Liu C, Tang YT;
XX WPI: 2001-639362/73.
DR P-PSDB; ABG20586.
XX New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.
XX Claim 1; SEQ ID NO 20577; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activities. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
SQ Sequence 2016 BP; 403 A; 327 C; 779 G; 501 T; 0 U; 0 Other;
Query March 18.7%; Score 231.4; DB 5; Length 2016;
Best Local Similarity 99.6%; Pred. No. 1.4e-43;
Matches 232; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 387 GAACCTGAATGACTTCATCCAGCTTGATGAAGAGTAGAGGTATGCAGAGTACCATT 446
DB 1141 GAACCTGAATGACTTCATCCAGCTTGATGAAGAGTAGAGGTATGCAGAGTACCATT 1200
QY 447 CTAGTTCGACGACCTGAGGAGACGACGAGCTGGCTAGTACCAGCAG 506
DB 1201 CTAGTTCGACGACCTGAGGAGACGACGAGCTGGCTAGTACCAGCAG 1260
QY 507 CAGTCTCAGGGCTCTGCCCAAGTACACGAGGACTACAGCTTCTGAACTGTGAGACAG 566
DB 1261 CAGTCTCAGGGCTCTGCCCAAGTACACGAGGACTACAGCTTCTGAACTGTGAGACAG 1320
QY 567 TCAGAGGCCACCAAGTAAGACTGCGTCTGACAAACGACCAAGTATGG 619

DB 1321 TCAGAGGCCACCAAGTAAGACTGCGTCTGACAAACGACCAAGTATGG 1373
RESULT 8
ABX63550
ID ABX63550 standard; cDNA; 1600 BP.
XX AC ABX63550;
XX 26-FEB-2003 (first entry)
DE Human cDNA #550 differentially expressed in activated vascular tissue.
XX Human; gene; ss; vascular tissue; cytostatic; atherosclerosis; cardiatic; hypertensive; antidiabetic; gynaecological; vasotropic; cerebroprotective; gene therapy; vascular disease; cancer; coronary; artery disease; hypertension; diabetes; pre-eclampsia; restenosis;
KW ischaemia-reperfusion injury; stroke.
XX Homo sapiens.
OS US2002137081-A1.
PN 26-SEP-2002.
XX 08-JAN-2002; 2002US-00044090.
XX 28-JUL-2000; 2000US-0222469P.
PR 08-JAN-2001; 2001US-0260483P.
XX (BAND/) BANDMAN O.
XX Bandman O;
XX WPI: 2003-110597/10.
XX Combination for diagnosing, staging, treating, or monitoring the progression of treatment of a vascular disease, e.g. atherosclerosis, comprises several cDNAs that are differentially expressed in activated vascular tissue.
XX Claim 1; Page; 18pp; English.
XX This invention relates to a combination comprising several cDNAs that are differentially expressed in activated vascular tissue. The invention also discloses a high throughput method for detecting differentially expressed cDNAs in a sample. The cDNAs of the invention may have antiatherosclerotic; cytostatic; cardiatic; hypertensive; antidiabetic; gynaecological; vasotropic and cerebroprotective activities and may be used in gene therapy. The cDNAs of the invention may be used in a high-throughput methods for detecting differential expression of one or more cDNAs in a sample, or screening several molecules or compounds to identify a molecule or compound that specifically binds a cDNA of the invention. A protein encoded by the cDNA may be used to screen several molecules or compounds to identify a ligand that specifically binds to the protein, or to produce or purify an antibody to the protein that can be used to detect a protein in a sample or purify a natural or recombinant protein from a sample. The nucleotides may be useful for diagnosing, staging, treating, or monitoring the progression of treatment of a vascular disease, e.g. atherosclerosis, cancer, coronary artery disease, hypertension, diabetes, pre-eclampsia, ischaemia-reperfusion injury, restenosis, or stroke. The cDNAs can also be used for large-scale genetic or gene expression analysis of several new nucleic acid molecules. Antibodies to the proteins encoded by the cDNAs are useful for diagnosing pre-pathologic disorders, and chronic or acute diseases associated with abnormalities in the expression, amount or distribution of the protein. The present sequence represents a cDNA of the invention that is differentially expressed in activated vascular tissue. Note: The sequence data for this patent did not form part of the specification, but was obtained in electronic format directly from USPTO at http.segdata.uspto.gov/sequence.html?docID=20020137081


```
XX AC ABX63552;
XX DT 26-FEB-2003 (first entry)
XX DE Human cDNA #552 differentially expressed in activated vascular tissue.
XX KW Human; gene; ss; vascular tissue; cytostatic; atherosclerosis; cardiant;
XX KW hypotensive; antidiabetic; gynaecological; vasotropic; cerebroprotective;
XX KW gene therapy; vascular disease; cancer; coronary; artery disease;
XX KW hypertension; diabetes; pre-eclampsia; restenosis;
XX KW ischaemia-reperfusion injury; stroke.
XX OS Homo sapiens.
XX PN US2002137081-A1.
XX PD 26-SEP-2002.
XX PF 08-JAN-2002; 2002US-00044090.
XX PR 28-JUL-2000; 2000US-0222469P.
XX PR 08-JAN-2001; 2001US-0260483P.
XX PA (BAND/) BANDMAN O.
XX PI Bandman O;
XX DR WPI; 2003-110597/10.
XX PT Combination for diagnosing, staging, treating, or monitoring the
XX PT progression of treatment of a vascular disease, e.g. atherosclerosis,
XX PT comprises several cDNAs that are differentially expressed in activated
XX PT vascular tissue.
XX PS Claim 1; Page; 18pp; English.
XX CC This invention relates to a combination comprising several cDNAs that are
XX CC differentially expressed in activated vascular tissue. The invention also
XX CC discloses a high throughput method for detecting differentially expressed
XX CC cDNAs in a sample. The cDNAs of the invention may have
XX CC antiatherosclerotic; cytostatic; cardiant; hypotensive; antidiabetic;
XX CC gynaecological; vasotropic and cerebroprotective activities and may be
XX CC used in gene therapy. The cDNAs of the invention may be used in a high-
XX CC throughput methods for detecting differential expression of one or more
XX CC cDNAs in a sample, or screening several molecules or compounds to
XX CC identify a molecule or compound that specifically binds a cDNA of the
XX CC invention. A protein encoded by the cDNA may be used to screen several
XX CC molecules or compounds to identify a ligand that specifically binds to
XX CC the protein, or to produce or purify an antibody to the protein that can
XX CC be used to detect a protein in a sample or purify a natural or
XX CC recombinant protein from a sample. The nucleotides may be useful for
XX CC diagnosing, staging, treating, or monitoring the progression of treatment
XX CC of a vascular disease, e.g. atherosclerosis, cancer, coronary artery
XX CC disease, hypertension, diabetes, pre-eclampsia, ischaemia-reperfusion
XX CC injury, restenosis, or stroke. The cDNAs can also be used for large-scale
XX CC genetic or gene expression analysis of several new nucleic acid
XX CC molecules. Antibodies to the proteins encoded by the cDNAs are useful for
XX CC diagnosing pre-pathologic disorders, and chronic or acute diseases
XX CC associated with abnormalities in the expression, amount or distribution
XX CC of the protein. The present sequence represents a cDNA of the invention
XX CC that is differentially expressed in activated vascular tissue. Note: The
XX CC sequence data for this patent did not form part of the specification, but
XX CC was obtained in electronic format directly from USPTO at
XX CC http://seqdata.uspto.gov/sequence.html?DocID=20020137081
XX CC Sequence 1691 BP; 514 A; 334 C; 395 G; 448 T; 0 U; 0 Other;
XX CC Query Match 18.3%; Score 226.2; DB 8; Length 1691;
XX CC Best Local Similarity 96.7%; Pred. No. 2.1e-42;
XX CC Matches 231; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
XX CC 1 ACATCCTTGGAATGCGACTAGCAACCAAGGGACAAGATATTCAAGAGTGTACTACTCAA 60
XX CC
```

409 ACAATCTTGTATGCGACTAGCAACCAAGGGACAAGATATTCAAGAGTGTACTACTCAA 468

61 TCCAGTACTCAAGCAAGTCCAGCAGCTCAGCGTTGCGCAACTTAGATCAACATGTTAG 120

469 TCCAGTACTCAAGCAAGTCCAGCAGCTCAGCGTTGCGCAACTTAGATCAACATGTTAG 528

121 ACCGAGCGATCAACTGTTTCTTAAATAAGAAAGTGAAGTGAAGCAAGCACTAAAGACA 180

529 ACCGAGCGATCAACTGTTTCTTAAATAAGAAAGTGAAGTGAAGCAAGCACTAAAGACA 588

181 AACTGGAACAAGCCCAAAATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 239

589 AACTGGAACAAGCCCAAAATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 647

RESULT 11
ABX63551/c
ID ABX63551 standard; cDNA; 1996 BP.
XX AC ABX63551;
XX DT 26-FEB-2003 (first entry)
XX DE Human cDNA #551 differentially expressed in activated vascular tissue.
XX KW Human; gene; ss; vascular tissue; cytostatic; atherosclerosis; cardiant;
XX KW hypotensive; antidiabetic; gynaecological; vasotropic; cerebroprotective;
XX KW gene therapy; vascular disease; cancer; coronary; artery disease;
XX KW hypertension; diabetes; pre-eclampsia; restenosis;
XX KW ischaemia-reperfusion injury; stroke.
XX OS Homo sapiens.
XX PN US2002137081-A1.
XX PD 26-SEP-2002.
XX PF 08-JAN-2002; 2002US-00044090.
XX PR 28-JUL-2000; 2000US-0222469P.
XX PR 08-JAN-2001; 2001US-0260483P.
XX PA (BAND/) BANDMAN O.
XX PI Bandman O;
XX DR WPI; 2003-110597/10.
XX PT Combination for diagnosing, staging, treating, or monitoring the
XX PT progression of treatment of a vascular disease, e.g. atherosclerosis,
XX PT comprises several cDNAs that are differentially expressed in activated
XX PT vascular tissue.
XX PS Claim 1; Page; 18pp; English.
XX CC This invention relates to a combination comprising several cDNAs that are
XX CC differentially expressed in activated vascular tissue. The invention also
XX CC discloses a high throughput method for detecting differentially expressed
XX CC cDNAs in a sample. The cDNAs of the invention may have
XX CC antiatherosclerotic; cytostatic; cardiant; hypotensive; antidiabetic;
XX CC gynaecological; vasotropic and cerebroprotective activities and may be
XX CC used in gene therapy. The cDNAs of the invention may be used in a high-
XX CC throughput methods for detecting differential expression of one or more
XX CC cDNAs in a sample, or screening several molecules or compounds to
XX CC identify a molecule or compound that specifically binds a cDNA of the
XX CC invention. A protein encoded by the cDNA may be used to screen several
XX CC molecules or compounds to identify a ligand that specifically binds to
XX CC the protein, or to produce or purify an antibody to the protein that can
XX CC be used to detect a protein in a sample or purify a natural or
XX CC recombinant protein from a sample. The nucleotides may be useful for
XX CC diagnosing, staging, treating, or monitoring the progression of treatment
XX CC of a vascular disease, e.g. atherosclerosis, cancer, coronary artery
XX CC disease, hypertension, diabetes, pre-eclampsia, ischaemia-reperfusion
XX CC injury, restenosis, or stroke. The cDNAs can also be used for large-scale
XX CC genetic or gene expression analysis of several new nucleic acid
XX CC molecules. Antibodies to the proteins encoded by the cDNAs are useful for
XX CC diagnosing pre-pathologic disorders, and chronic or acute diseases
XX CC associated with abnormalities in the expression, amount or distribution
XX CC of the protein. The present sequence represents a cDNA of the invention
XX CC that is differentially expressed in activated vascular tissue. Note: The
XX CC sequence data for this patent did not form part of the specification, but
XX CC was obtained in electronic format directly from USPTO at
XX CC http://seqdata.uspto.gov/sequence.html?DocID=20020137081
XX CC Sequence 1691 BP; 514 A; 334 C; 395 G; 448 T; 0 U; 0 Other;

QY 1195 TACTTTGAATGTTGCAAAAGGGCAATTT 1222
Db 31 TACTTTGAATGTTGCTAAAAGGACATTT 4

Search completed: November 14, 2004, 13:05:30
Job time : 687 secs

XX 06-NOV-2001 (first entry)
XX Human polynucleotide SEQ ID NO 4810.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
XX WO200164835-A2.
XX
XX 07-SEP-2001.
XX
XX 26-FEB-2001; 2001WO-US004927.
XX
XX 28-FEB-2000; 2000US-00515126.
XX
XX 18-MAY-2000; 2000US-00577409.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
XX
XX WPI; 2001-514838/56.
XX
XX P-PSDB; AAO04819.
XX
XX Isolated nucleic acids and polypeptides, useful for preventing diagnosing
XX and treating e.g. leukemia, inflammation and immune disorders.
XX
XX Claim 1; SEQ ID NO 4810; 1399pp + Sequence Listing; English.
XX
XX The invention relates to human polynucleotides (AA179941-AA193841) and
XX the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation. Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pat_sequences
XX
SQ Sequence 330 BP; 151 A; 63 C; 49 G; 67 T; 0 U; 0 Other;

Query Match 13.2%; Score 163.6; DB 4; Length 330;
Best Local Similarity 70.7%; Pred. No. 4.3e-28;
Matches 232; Conservative 0; Mismatches 94; Indels 2; Gaps 1;

QY 895 GTAATCGAAGTGTGGGTCCGGCCAGCTTCAGAAATGCTTGACTCAAGTGAATGTAC 954
Db 329 GAAATTAACCTGGGTTCGCCCCCTTTAAAGAGGTGGGCCCAAGTTAAAAACC 270

QY 955 AGGTTTCAGTTTGTATATTTTTCAGCAATTTTATACAGTGTCAATTTGGGA 1014
Db 269 CGGGGCCCGTTTTTAAAAATTTTTCGCCCAATTTTTCAGGGTCTTTTTTTGGGG 210

QY 1015 GAGGATCTGTCAGAAAAATTAATGATACATTTTGTCAAAATTTGCTTTTGTGGGTG 1074
Db 209 AAGGAATCTTTCCAAAAAAGAAAAATTTTTTCCCAATTCCTTTTGGGGGG 150

QY 1075 ACCTTTTGGTTTTTTTGTGTGTTTTTCTTTGGATTTTTTCTTTCTTTTAAAG 1134
Db 149 GGGGTTGGGTTTTTTTGGGGTTTTTTTGGGTTTTTTTGGGTTTTTTT--T 92

QY 1135 AGGTGTTGTGTTTTTTTTCCTTCAATACATCTCGCGGCTTGGGAATGTAACAGTTAAT 1194
Db 91 TTTTTTTTTTTTTTTTGTGCTTCATCTCTCGCGCTTGGAAATGTACAGTTAAT 32

Result No.	Query \$			DB	ID	Description
	Score	Match	Length			
1	1238	100.0	1238	6	AX069004	Sequence
2	1174.6	94.9	2108	9	BC069192	BC069192 Homo sapi
3	1159.8	93.7	2134	9	AF374416	AF374416 Homo sapi
4	980	79.2	1802	10	MWU276707	AX276707 Mus muscu
5	953.8	77.0	2291	6	CQ845798	CQ845798 Sequence
6	943.8	77.0	2291	9	AK131393	AK131393 Homo sapi
7	943.8	76.2	1167	9	HSA276706	AX276706 Homo sapi
8	797.8	64.4	165802	9	AL135914	AL135914 Human DNA
9	666.6	53.8	203077	10	AC127172	AC127172 Mus muscu
10	666.6	53.8	203949	10	ALU589878	ALU589878 Mouse DNA
11	654.6	52.9	110000	2	AC074222.1	Continuation (2 of
12	654.6	52.9	110000	2	AC074222.2	Continuation (3 of
13	654	52.8	1888	5	BX950648	BX950648 Gallus ga
14	608.6	49.2	165096	2	AC112423	AC112423 Rattus no
15	608.6	49.2	246585	2	AC108954	AC108954 Rattus no
16	590.8	47.7	2179	5	BC063362	BC063362 Xenopus t
17	561.8	45.4	177389	10	AC123453	AC123453 Mus muscu
18	403.8	32.6	1936	5	BC055544	BC055544 Danio rer
19	227.8	18.4	3733	9	AF277190	AF277190 Homo sapi

Qy	241	GGAAAAAGTTAATGCGAAGTGTGAAATGCTTATCCAGGAGAAATCAAGAGCTTGGAGGC	300
Ds	241	GGAAAAAGTTAATGCGAAGTGTGAAATGCTTATCCAGGAGAAATCAAGAGCTTGGAGGC	300
Qy	301	AGCTCTCCAGGAGCTATTGCAACATTGGAAGCAGAGTGGCTTTACAGAGAAATAC	360
Ds	301	AGCTCTCCAGGAGCTATTGCAACATTGGAAGCAGAGTGGCTTTACAGAGAAATAC	360
Qy	361	GTGAGAGCTTAAGAGCAGTTCAGATGAATGACTTATCATCCAGCTTGATCAAG	420
Ds	361	GTGAGAGCTTAAGAGCAGTTCAGATGAATGACTTATCATCCAGCTTGATCAAG	420
Qy	421	AAGTAGAGGGTATGACAGTAGTACCATTTCTGACAGCAGCTGAAGGAGACACGCC	480
Ds	421	AAGTAGAGGGTATGACAGTAGTACCATTTCTGACAGCAGCTGAAGGAGACACGCC	480
Qy	481	AGCAGTTGGCTCAGTACAGCAGCAGTCTCAGGCTCTGCCCAAGTACACGAGGA	540
Ds	481	AGCAGTTGGCTCAGTACAGCAGCAGTCTCAGGCTCTGCCCAAGTACACGAGGA	540
Qy	541	CTACAGCTTCTGAACCTGTAGAACAGTCCAGAGGCCAACAGTAAGACTGCAGTCTGA	600
Ds	541	CTACAGCTTCTGAACCTGTAGAACAGTCCAGAGGCCAACAGTAAGACTGCAGTCTGA	600
Qy	601	CAACGGACCAAGTAATGTTAGTCTCTCCGCCAGAGAGCGTCTGGGTCTGGATTACA	660
Ds	601	CAACGGACCAAGTAATGTTAGTCTCTCCGCCAGAGAGCGTCTGGGTCTGGATTACA	660
Qy	661	GGGAGGGCAACACAAACCAAGATCACTTCTCTCCAGGGAATGTAATAAGTCCT	720
Ds	661	GGGAGGGCAACACAAACCAAGATCACTTCTCTCCAGGGAATGTAATAAGTCCT	720
Qy	721	CCACAGCTCAGAGAGAGAACTGGCAGAGAGTAGTGTACGTAAATCAACTCAGTG	780
Ds	721	CCACAGCTCAGAGAGAGAACTGGCAGAGAGTAGTGTACGTAAATCAACTCAGTG	780
Qy	781	CGGGGTGAAGTGAAGTCTCCACGGGAGTGAAGTCTCTCACACCAATCAA	840
Ds	781	CGGGGTGAAGTGAAGTCTCCACGGGAGTGAAGTCTCTCACACCAATCAA	840
Qy	841	ATGACACAGACTCCAGTCAATGCTCAAGAGGAGAAAGCAGTCAAGTGAAGTAAATC	900
Ds	841	ATGACACAGACTCCAGTCAATGCTCAAGAGGAGAAAGCAGTCAAGTGAAGTAAATC	900
Qy	901	GAACTGGGTTCCGCCAGCTTCAAGATGGCTGGACTCAAGTGAAGTCAAGGTT	960
Ds	901	GAACTGGGTTCCGCCAGCTTCAAGATGGCTGGACTCAAGTGAAGTCAAGGTT	960
Qy	961	CAGTTTGTAAATATTTTTCAGCAAAATTTTATACAGTGCATTTAAATTTGGAGAGGAT	1020
Ds	961	CAGTTTGTAAATATTTTTCAGCAAAATTTTATACAGTGCATTTAAATTTGGAGAGGAT	1020
Qy	1021	ACTCTCAGAAATTAATGATCACTTTTGTGCAATTTGCTTTTGTGGGTGACGTT	1080
Ds	1021	ACTCTCAGAAATTAATGATCACTTTTGTGCAATTTGCTTTTGTGGGTGACGTT	1080
Qy	1081	TGGTTTTTTTGTGTTTTTTTTCAGCAAAATTTTATACAGTGCATTTAAATTTGGAGAGGAT	1140
Ds	1081	TGGTTTTTTTGTGTTTTTTTTCAGCAAAATTTTATACAGTGCATTTAAATTTGGAGAGGAT	1140
Qy	1141	TGTGTTTTTTTGTGTTTTTTTTCAGCAAAATTTTATACAGTGCATTTAAATTTGGAGAGGAT	1200
Ds	1141	TGTGTTTTTTTGTGTTTTTTTTCAGCAAAATTTTATACAGTGCATTTAAATTTGGAGAGGAT	1200
Qy	1201	GAAATGTCGAAAGGGGATTTTGTGGAGGTCAGTTA	1238
Ds	1201	GAAATGTCGAAAGGGGATTTTGTGGAGGTCAGTTA	1238

RESULT 2
BC069192
LOCUS

DEFINITION	Homo sapiens Wilms tumor 1 associated protein, transcript variant 1, mRNA (CDNA clone MGC:78409 IMAGE:4653831), complete cds.
ACCESSION	BC069192
VERSION	BC069192.1 GI:46623319
KEYWORDS	MGC.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	1 (bases 1 to 2108) Mammalia: Eutheria; Primates; Catarrhini; Hominidae; Homo. Straussberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diachenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Udwin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahy, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Smalls, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
PUBLISHED	12477932
REFERENCE	2 (bases 1 to 2108) Straussberg, R.
AUTHORS	Direct Submission
TITLE	Submitted (26-Apr-2004) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov Contact: MGC help desk Email: cgapbs-r@mail.nih.gov Tissue procurement: ATCC CDNA Library Preparation: Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada info@bcgsc.bc.ca Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butterfield, Andy Chan, Steve S. Chand, William Chow, Alison Cloutier, Ruth Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao, Kim MacDonald, Anara Masson, Mike R. Mayo, Josh Moran, Ryan Morin, Teika Olson, Diana Palmquist, Anca Petrescu, Anna Liisa Prabu, Parvaneh Saeedi, JR Santos, Angelique Schnerch, Ursula Skalska, Duane Smalls, Jeff Stott, Miranda Tsai, George Yang, Jacquie Schein, Asim Siddiqui, Rob Holt, Marco Marra.
COMMENT	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAL Plate: 54 Row: d Column: 11 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 23199975.
FEATURES	Location/Qualifiers 1..2108 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="MGC:78409 IMAGE:4653831" /tissue_type="placenta, chorioncarcinoma" /lab_host="NIH MGC 21" /lab_host="DH10B-R" /note="Vector: pOTB7"
source	1..2108
gene	

Query Match 94.9%; Score 1174.6; DB 9; Length 2108;
Best Local Similarity 97.8%; Pred. No. 3.4e-256;
Matches 1212; Conservative 0; Mismatches 24; Indels 3; Gaps 2;

QY 1 ACATCCTTGAATGCGACTAGCAACCAAGGAGCAAGATATTCAGAGTGCTACTACAA 60
DB 401 ACATCCTTGAATGCGACTAGCAACCAAGGAGCAAGATATTCAGAGTGCTACTACAA 460
QY 61 TCCAGTACCTCAAGCAAGTCAGCAGCCTAGCGTTGCCAACTTAGATCAACAATGGTAG 120
DB 461 TCCAGTACCTCAAGCAAGTCAGCAGCCTAGCGTTGCCAACTTAGATCAACAATGGTAG 520
QY 121 ACCAGAGATCACTTGTCTTCTTAAATGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 180
DB 521 ACCAGAGATCACTTGTCTTCTTAAATGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 580
QY 181 AACTGGCAACAGCCCAAAATGAATGAGTGCCTGGAAGTGTAGCGCTGTAGTACCAACAG 240
DB 581 AACTGGCAACAGCCCAAAATGAATGAGTGCCTGGAAGTGTAGCGCTGTAGTACCAACAG 640
QY 241 GGAAAAAGTTAATGGCAAGTGTGCAATGCTTATCCAGGAGATCAAGAGCTTGGAGGC 300
DB 641 GGAAAAAGTTAATGGCAAGTGTGCAATGCTTATCCAGGAGATCAAGAGCTTGGAGGC 700
QY 301 AGCTGTCCAGGAGGATTCACCAACTTGAAGCAGAGTGGCTTTTACAGAGAAATACA 360
DB 701 AGCTGTCCAGGAGGATTCACCAACTTGAAGCAGAGTGGCTTTTACAGAGAAATACA 760
QY 361 GTGAGGAGCTTAAAGCAGTCAGGATGAATGATCTTCAATCATCCAGCTTGATGAAG 420
DB 761 GTGAGGAGCTTAAAGCAGTCAGGATGAATGATCTTCAATCATCCAGCTTGATGAAG 820
QY 421 AAGTAGAGGTATGAGAGTACCATCTAGTCTGACAGCAGCTGAAGGAGACAGCC 480
DB 821 AAGTAGAGGTATGAGAGTACCATCTAGTCTGACAGCAGCTGAAGGAGACAGCC 880
QY 481 AGCAGTTGGCTCAGTACCAAGCAGCAGTCTAGCGCTCTGCCCCCAAGTACCAAGCAGGA 540
DB 881 AGCAGTTGGCTCAGTACCAAGCAGCAGTCTAGCGCTCTGCCCCCAAGTACCAAGCAGGA 940
QY 541 CTACAGCTTCTGAACCTGTAGAACAGTCAAGGAGCACAAGTAAAGACTGCACTGCTGA 600
DB 941 CTACAGCTTCTGAACCTGTAGAACAGTCAAGGAGCACAAGTAAAGACTGCACTGCTGA 1000
QY 601 CAAACGGACCAAGTAAATGGTAGCTCTCCCGCAGAGGAGCTCTGGGTCTGGATTTACCA 660
DB 1001 CAAACGGACCAAGTAAATGGTAGCTCTCCCGCAGAGGAGCTCTGGGTCTGGATTTACCA 1060
QY 661 GGGAGGGCAACACACCGAAGATGACTTCTTCTTCTCCAGGAGAAAGTGAATGAGTCT 720
DB 1061 GGGAGGGCAACACACCGAAGATGACTTCTTCTTCTTCTCCAGGAGAAAGTGAATGAGTCT 1120

ORIGIN

Query Match 94.9%; Score 1174.6; DB 9; Length 2108;
Best Local Similarity 97.8%; Pred. No. 3.4e-256;
Matches 1212; Conservative 0; Mismatches 24; Indels 3; Gaps 2;

QY 1 ACATCCTTGAATGCGACTAGCAACCAAGGAGCAAGATATTCAGAGTGCTACTACAA 60
DB 401 ACATCCTTGAATGCGACTAGCAACCAAGGAGCAAGATATTCAGAGTGCTACTACAA 460
QY 61 TCCAGTACCTCAAGCAAGTCAGCAGCCTAGCGTTGCCAACTTAGATCAACAATGGTAG 120
DB 461 TCCAGTACCTCAAGCAAGTCAGCAGCCTAGCGTTGCCAACTTAGATCAACAATGGTAG 520
QY 121 ACCAGAGATCACTTGTCTTCTTAAATGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 180
DB 521 ACCAGAGATCACTTGTCTTCTTAAATGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 580
QY 181 AACTGGCAACAGCCCAAAATGAATGAGTGCCTGGAAGTGTAGCGCTGTAGTACCAACAG 240
DB 581 AACTGGCAACAGCCCAAAATGAATGAGTGCCTGGAAGTGTAGCGCTGTAGTACCAACAG 640
QY 241 GGAAAAAGTTAATGGCAAGTGTGCAATGCTTATCCAGGAGATCAAGAGCTTGGAGGC 300
DB 641 GGAAAAAGTTAATGGCAAGTGTGCAATGCTTATCCAGGAGATCAAGAGCTTGGAGGC 700
QY 301 AGCTGTCCAGGAGGATTCACCAACTTGAAGCAGAGTGGCTTTTACAGAGAAATACA 360
DB 701 AGCTGTCCAGGAGGATTCACCAACTTGAAGCAGAGTGGCTTTTACAGAGAAATACA 760
QY 361 GTGAGGAGCTTAAAGCAGTCAGGATGAATGATCTTCAATCATCCAGCTTGATGAAG 420
DB 761 GTGAGGAGCTTAAAGCAGTCAGGATGAATGATCTTCAATCATCCAGCTTGATGAAG 820
QY 421 AAGTAGAGGTATGAGAGTACCATCTAGTCTGACAGCAGCTGAAGGAGACAGCC 480
DB 821 AAGTAGAGGTATGAGAGTACCATCTAGTCTGACAGCAGCTGAAGGAGACAGCC 880
QY 481 AGCAGTTGGCTCAGTACCAAGCAGCAGTCTAGCGCTCTGCCCCCAAGTACCAAGCAGGA 540
DB 881 AGCAGTTGGCTCAGTACCAAGCAGCAGTCTAGCGCTCTGCCCCCAAGTACCAAGCAGGA 940
QY 541 CTACAGCTTCTGAACCTGTAGAACAGTCAAGGAGCACAAGTAAAGACTGCACTGCTGA 600
DB 941 CTACAGCTTCTGAACCTGTAGAACAGTCAAGGAGCACAAGTAAAGACTGCACTGCTGA 1000
QY 601 CAAACGGACCAAGTAAATGGTAGCTCTCCCGCAGAGGAGCTCTGGGTCTGGATTTACCA 660
DB 1001 CAAACGGACCAAGTAAATGGTAGCTCTCCCGCAGAGGAGCTCTGGGTCTGGATTTACCA 1060
QY 661 GGGAGGGCAACACACCGAAGATGACTTCTTCTTCTCCAGGAGAAAGTGAATGAGTCT 720
DB 1061 GGGAGGGCAACACACCGAAGATGACTTCTTCTTCTTCTCCAGGAGAAAGTGAATGAGTCT 1120

QY 721 CCAACAGCTCAGAGGAGAGACTGGCAGAGAGAGTGTAGTGTACGTAATCAACTCAGTG 780
DB 1121 CCAACAGCTCAGAGGAGAGAGACTGGCAGAGAGAGTGTAGTGTACGTAATCAACTCAGTG 1180
QY 781 CCGGGTATGAAAGTGTAGACTCTCCACGGGCAAGTGAATACTCTCTCACACCAATCAA 840
DB 1181 CCGGGTATGAAAGTGTAGACTCTCCACGGGCAAGTGAATACTCTCTCACACCAATCAA 1240
QY 841 ATGACACAGACTCCAGTCATGACCTCAAGAGGAGAGAGCAGTGTGGGAAAGTAAATC 900
DB 1241 ATGACACAGACTCCAGTCATGACCTCAAGAGGAGAGAGCAGTGTGGGAAAGTAAATC 1300
QY 901 GAACTGTGGGTTCGCCGCCACAGCTTCAGATGCTTGGACTCAAGTGAATCAACAGGTT 960
DB 1301 GAACTGTGGGTTCGCCGCCACAGCTTCAGATGCTTGGACTCAAGTGAATCAACAGGTT 1360
QY 961 CAGTTTTGTATATATTTTTTTCAGCAAAATTTTATACAGTGTCAATTAATTTGGGAGAGAT 1020
DB 1361 CAGTTTTGTATATATTTTTTTCAGCAAAATTTTATACAGTGTCAATTAATTTGGGAGAGAT 1420
QY 1021 ACTGTCCAGAGAAATTAATGATATCTTTTGTGCATATTTTGTGCATATTTTGTGGGTGTACGTTT 1080
DB 1421 ACTGTCCAGAGAAATTAATGATATCTTTTGTGCATATTTTGTGGGTGTACGTTT 1480
QY 1081 TGGTTTTTTTTTGT 1140
DB 1481 TGGTTTTTTTTTGT 1538
QY 1141 TTGTTGT 1200
DB 1539 TTTTTTTTTTTTGT 1598
QY 1201 GAATGTTGC-AAAAGGCAATTTTGTGGAGGTTCCAGTTA 1238
DB 1599 GAATGTTGTCTAAAGACATTTTGTAGGGTCAAGTTA 1637

RESULT 3
AF374416 2134 bp mRNA linear PRI 21-MAY-2001
LOCUS
DEFINITION Homo sapiens putative pre-mRNA splicing regulator female-lethal (2D)
ACCESSION AF374416
VERSION AF374416.1 GI:14161494
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini, Hominidae; Homo.
REFERENCE 1 (bases 1 to 2134)
AUTHORS Ortega A., Penalba, L. and Valcarcel, J.
TITLE cDNA of human homolog of the drosophila putative pre-mRNA splicing regulator female-lethal (2D) [FL(2)D]
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2134)
AUTHORS Ortega A., Penalba, L. and Valcarcel, J.
TITLE Direct Submission
JOURNAL Submitted (26-APR-2001) Gene Expression Programme, European Molecular Biology Laboratory, Meyerhofstrasse 1, Heidelberg, Baden-Württemberg 69117, Germany
FEATURES
Location/Qualifiers
1..2134
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
226..1416
/chromosome="6"
/note="FL(2)D; similar to Homo sapiens WTAP protein encoded by GenBank Accession Number AJ276706"
/codon_start=1
/product="putative pre-mRNA splicing regulator female-lethal (2D)"

FFLMKGELEOTKDLQEQNELSANKFTPDSTQTKYKLMKCRMLIOENOEJGROLSQ
GRIQAEALUOKYSEELKXSODELNDPILCLDEVEGMSSTILVLOQKXETREQ
LAQYQOQOASAPSTSTSEPSVDOAEVTKDCKRLANGSPNSGSGGFFH
REGSTPDDPSSNGNXXSNSEBRTGSGSSYIPLSAGYVESYDPSGSENSLTH
HSNDTDSHDPQBEKAVSGKGNRTVGRSHVQNGLDSSVNVQAVL"

ORIGIN		Query Match		79.28; Score 980; DB 10; Length 1802;	
		Best Local Similarity		89.2%; Pred. No. 5.5e-212;	
		Matches 1105; Conservative		0; Mismatches 96; Indels 38; Gaps 3;	
Qy	1	ACATCCCTTGAATGCGACTAGCAACCAAGGAGCAAGATATTCAGAGGTGCTACTCAAA	60		
Db	170	ACATCTTGTATGTCGCGCTAGCAACCAAGGAGGAGATGCAAGAGTGCAACACTCAA	229		
Qy	61	TCCAGTACCTCAAGCAAGTCCAGAGCCTAGCGTTGCCCACTTAGATCAACAATGGTAG	120		
Db	230	TCCAGTACCTCAAGCAAGTCCAGAGCCTAGCGTTGCCCACTTAGATCAACAATGGTAG	289		
Qy	121	ACCAGCGATCAACTCTGTTTCTCTAAATGAAAGGTGAACCTGGAACAGACTAAAGACA	180		
Db	290	ACCAGCGATCAACTCTGTTTCTCTAAATGAAAGGTGAACCTGGAACAGACTAAAGACA	349		
Qy	181	AACTGGCAAGCCCAAAATGAATGAGTGCCTGGAAAGTTTACGCTGATAGCCAAACAG	240		
Db	350	AACTGGCAAGCCCAAAATGAATGAGTGCCTGGAAAGTTTACGCTGATAGCCAAACAG	409		
Qy	241	GGAAGAGTTAATGCGAGGTGTCGAATGCTTATCCAGGAGATCAAGAGCTTGGAGGC	300		
Db	410	GCAAAAAGCTAATGGCAAGTGTGCAATGCTTATCCAGGAGAACCAAGAGCTTGGAGGC	459		
Qy	301	AGCTGTCCCGAGGAGCTATTGCAACACTTGAAGCAGAGTTGGCTTTACAGAAATACA	360		
Db	470	AGCTGTCCCGAGGAGCTATTGCAACACTTGAAGCAGAGTTGGCTTTACAGAAATACA	529		
Qy	361	GTGAGGAGCTTAAAGAGCTCAGATGAATGATCTCATCATCCAGCTTGATGAAG	420		
Db	530	GTGAGGAGCTTAAAGAGCTCAGATGAATGATCTCATCATCCAGCTTGATGAAG	589		
Qy	421	AAAGTAGGGTATGACAGAGTACCTTCTAGTTCTGACAGCAGCTGAAGGAGACGCC	480		
Db	590	AAAGTAGGGTATGACAGAGTACCTTCTAGTTCTGACAGCAGCTGAAGGAGACGCC	649		
Qy	481	AGCAGTTGGCTCAGTACAGCAGAGCTCAGGCTCTGCCCCCAAGTACAGCAGGA	540		
Db	650	AGCAGTTGGCTCAGTACAGCAGAGCTCAGGCTCTGCCCCCAAGTACAGCAGGA	709		
Qy	541	CTACAGCTTCTGAACCTGTAGACAGTACAGAGCCACAAAGTAAAGACTGCTGCTGA	600		
Db	710	CTACATCTTCTGAACCTGTAGATCAGGAGAGGTCACAGCAAGAGCTGCACTGCTGG	769		
Qy	601	CAACGGACCAAGTAAATGGTAGCTCTCCCGCAGAGAGCTCTGGGTCTGGATTACA	660		
Db	770	CAATGGACCAAGTAAATGGTAGCTCTCCCGCAGAGAGCTCTGGGTCTGGATTACA	829		
Qy	661	GGGAGGCAACACACACCAAGTACTTCTCTCTCCAGGGAATGGTAATAGTCT	720		
Db	830	GGGAGGAGGACACCTCAGGATGACTTCTCTCTCTCTCAGGGAATGGTAATAGTCT	889		
Qy	721	CCAAAGCTCAGAGGAGAACTGGCAGAGAGGTAGTGGTTACGTAAATCAATCAGTG	780		
Db	890	CCAAAGCTCAGAGGAGAACTGGCAGAGAGGTAGTGGTTACGTAAATCAATCAGTG	949		
Qy	781	CGGGGTATGAAGTGTAGCTCTCCAGGGGAGTGAAACTCTCTCACACACCAATCAA	840		
Db	950	CGGGGTATGAAGTGTAGCTCTCCAGGGGAGTGAAACTCTCTCACACATCACTCAA	1009		
Qy	841	ATGACACAGCTCCAGTCAATGACCTCAAGAGGAGAGAGCAGTGAAGTGGGAAAGTATC	900		
Db	1010	ATGACACAGCTCCAGTCAATGACCTCAAGAGGAGAGAGCAGTGAAGTGGGAAAGTATC	1069		
Qy	901	GAACTGTGGGTTCCCGCCACAGTTTCAGATGGCTTGGACTCAAGTGTAAATGTACAGGTT	960		

Db	1070	GAACTGTGGGTTCCCGCCATGTTCCAGATGGCTTGGACTCAAGTGTAAATGTACAGGGTG	1129		
Qy	961	CAGTTTGTAAATATTTTTCAGCAAAATTTTATACAGTGTCAATTTAATTTGGGAGAGGAT	1020		
Db	1130	CAGTTTGTAAATATTTTTCAGCAAAATTTTATACAGTGTCAATTTAATTTGGGAGAGGGA	1189		
Qy	1021	ACTGTCCAGAAAATTAATGCATCTTTTGTGCACAATTTGCCCTTTTGTGGGTGACGTTT	1080		
Db	1190	TACTGTCAAAAATTAATGCATCTTTTGTGCACAATTTACCTTTTGTGGGTGT--GTTT	1247		
Qy	1081	TGTTTTTTTTTGTGTTTTTTTTTCTTTGGAATTTTTTTTTCTTTCTTTTAAAGAGGTGG	1140		
Db	1248	TTGTTTTTGTGTTGTTCTTTTTTTTTT-----	1275		
Qy	1141	TTGTTGTTTTTTTTGCTTCAATCTCTGCGCGCTTTGGGAATTTGTAAACAGTTAATTTACTTT	1200		
Db	1276	---TTTTTTTAATTGCCTTCAATACCTCTGCCACTTTGAAATTTGTAACAGTTAATTTACTTT	1332		
Qy	1201	GAAATGTGTC-AAAAGGGCATTTTGTGGAGGGTCCAGTTA	1238		
Db	1333	GAACATGTCTAAAGGACATTTTGTAGGGTCAAGTTA	1371		
RESULT 5					
CQ845798		2291 bp		DNA linear PAT 02-AUG-2004	
LOCUS		Sequence 4445 from Patent EP1440981.			
DEFINITION		CQ845798			
ACCESSION		CQ845798.1		GI:50895397	
VERSION					
KEYWORDS		Homo sapiens (human)			
SOURCE		Homo sapiens			
ORGANISM		Homo sapiens			
REFERENCE		1			
AUTHORS		Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S., Yamamoto,J., Isono,Y., Nagai,K. and Irie,R.			
TITLE		Full-length human cdna			
JOURNAL		Patent: EP 1440981-A 4445 28-JUL-2004;			
FEATURES		Research Association for Biotechnology (JP)			
source		Location/Qualifiers			
		1..2291			
		/organism="Homo sapiens"			
		/mol_type="unassigned DNA"			
		/db_xref="taxon:9606"			
ORIGIN					
Query Match		77.0%; Score 953.8; DB 6; Length 2291;			
Best Local Similarity		97.9%; Pred. No. 5e-206;			
Matches 988; Conservative		0; Mismatches 17; Indels 4; Gaps 2;			
Qy	231	AGCCAAACAGGGGAAAAGTTAATGGCGAAGTGTGCAATGCTTATCCAGGAGATCAAGAG	290		
Db	841	AGCCAAACAGGGGAAAAGTTAATGGCGAAGTGTGCAATGCTTATCCAGGAGATCAAGAG	900		
Qy	291	CTTGGGAAGCAGCTGTCCAGGAGAGCTATTGCACAACTTGAAGCAGAGTTGGCTTTACAG	350		
Db	901	CTTGGGAAGCAGCTGTCCAGGAGAGCTATTGCACAACTTGAAGCAGAGTTGGCTTTACAG	960		
Qy	351	AAGAAATACAGTCAGGAGCTTAAAGCAGTCAGGATGAATGAATGACTTCATCATCCAG	410		
Db	961	AAGAAATACAGTCAGGAGCTTAAAGCAGTCAGGATGAATGAATGACTTCATCATCCAG	1020		
Qy	411	CTTGTATGAAGAAGTAGAGGATGTCAGAGTACCATTCTAGTTCTGTGAGCAGCAGCTGAAG	470		
Db	1021	CTTGTATGAAGAAGTAGAGGATGTCAGAGTACCATTCTAGTTCTGTGAGCAGCAGCTGAAG	1080		
Qy	471	GAGACAGCCAGCAGTTGGCTCAGTACAGCAGCAGCAGCTCTGAGCCTCTGCCCCAAGT	530		
Db	1081	GAGACAGCCAGCAGTTGGCTCAGTACAGCAGCAGCAGCTCTGAGCCTCTGCCCCAAGT	1140		
Qy	531	ACCAGCAGGACTACAGCTTCTGAACCTGTGAAACAGTACAGAGCCCAAGTAAAGACTGCC	590		

```

Db      1141 ACCAGGAGGACTACAGCTCTCTGAACTGTAGACAGTCTAGAGGCCACAAAGTAAAGACTGC 1200
QY      591 AGTCGTCTGACAAACGGACCAAGTAAATGGTAGCTCTCTCCGCCAGAGGACGCTCTGGGTCT 650
Db      1201 AGTCGTCTGACAAACGGACCAAGTAAATGGTAGCTCTCTCCGCCAGAGGACGCTCTGGGTCT 1260
QY      651 GGATTTTCACAGGAGGGGCAACCAACCGAAGATGATTTCTCTCTCTCCAGGGAATGGT 710
Db      1261 GGATTTTCACAGGAGGGGCAACCAACCGAAGATGATTTCTCTCTCTCCAGGGAATGGT 1320
QY      711 AATAAGTCTTCAACAGCTCAGAGGAGAGAACTGCGCAGAGGAGGTAGTGGTTACGTAAAT 770
Db      1321 AATAAGTCTTCAACAGCTCAGAGGAGAGAACTGCGCAGAGGAGGTAGTGGTTACGTAAAT 1380
QY      771 CAACCTCAGTGGGGGTATGAAGTGTAGACTCTCCAGGGCAGTGAAGTCTCTCACA 830
Db      1381 CAACCTCAGTGGGGGTATGAAGTGTAGACTCTCCAGGGCAGTGAAGTCTCTCACA 1440
QY      831 CACCAATCAAAATGACACAGACTCCAGTCCATGACCTCAAGAGGAGAAAGCAGTGGG 890
Db      1441 CACCAATCAAAATGACACAGACTCCAGTCCATGACCTCAAGAGGAGAAAGCAGTGGG 1500
QY      891 AAAGTAATCGAATGCTGGGTTCGGGCCAGCTTCCAGAGTTCAGAGTGGTGAAGTGAAT 950
Db      1501 AAAGTAATCGAATGCTGGGTTCGGGCCAGCTTCCAGAGTTCAGAGTGGTGAAGTGAAT 1560
QY      951 GTACAGGGTTCAGTCTTGTATATTTTTCAGCAAAATTTTATACAGTGTCAATTAATTT 1010
Db      1561 GTACAGGGTTCAGTCTTGTATATTTTTCAGCAAAATTTTATACAGTGTCAATTAATTT 1620
QY      1011 GGGAGAGATGACTGTCAGAAAATTAATGACATCTTTGTCACAAATTTGCTTTTGTGG 1070
Db      1621 GGGAGAGATGACTGTCAGAAAATTAATGACATCTTTGTCACAAATTTGCTTTTGTGG 1680
QY      1071 GTGACGTTTGTGTTTTTTTGTGTTTTTTTGTGTTTTTTTGTGTTTTTTTGTGTTTTTT 1130
Db      1681 GTGACGTTTGTGTTTTTTTGTGTTTTTTTGTGTTTTTTTGTGTTTTTTTGTGTTTTTT 1740
QY      1131 TAAGAGTGGTGTGTTTTTTTGTGTTTTTTTGTGTTTTTTTGTGTTTTTTTGTGTTTTTT 1190
Db      1741 T---TTTTTTTTTTTTTTTTTGTCTTCAATCTCTGCGCTTTCGAAATTTGAACAGT 1797
QY      1191 TAATTACTTGAATGTTGC-AAAGGGCATTTGTGAGGGTCCAGTTA 1238
Db      1798 TAATTACTTGAATGTTGCTAAAGGACATTTTGTAGGGTCAAGTTA 1846

RESULT 6
AKI31393
LOCUS      2291 bp mRNA linear PRI 07-MAY-2004
DEFINITION Homo sapiens cDNA FLJ16480 fis, clone BRHA2016318, highly similar to WRAP protein.
ACCESSION AKI31393
VERSION   AKI31393.1 GI:47077246
KEYWORDS  oligo capping; fis (full insert sequence).
SOURCE    Homo sapiens
ORGANISM  Homo sapiens
REFERENCE 1
AUTHORS   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Ninomiya,K., Wagatsuma,M., Kanda,K., Kondo,H., Yokoi,T., Kodaira,H., Furuya,T., Takahashi,M., Kikkawa,E., Omura,Y., Abe,K., Kamihara,K., Katsuta,N., Sato,K., Tanikawa,M., Yamazaki,M., Sugiyama,T., Irie,Y., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Murakawa,K., Kanehori,K., Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahara,K., Masuho,Y., Nagai,K. and Isogai,T.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 2291)
Isogai,T. and Yamamoto,J.

TITLE
JOURNAL
REFERENCE
AUTHORS

```

TITLE JOURNAL

COMMENT

Direct Submission
Submitted (01-MAR-2004) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.

FEATURES

source

1..2291
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="BRHA2016318"
/tissue_type="thalamus"
/clone_lib="BRHA2"
/note="cloning vector: pME18SFL3"

ORIGIN

Query Match 77.0%; Score 953.8; DB 9; Length 2291;
Best Local Similarity 97.9%; Pred. No. Se-206;
Matches 988; Conservative 0; Mismatches 17; Indels 4; Gaps 2;

```

QY      231 ACCCAACAGGGAAGAAAGTTAATGCGAAGTGTGCAATGCTTATCCAGAGAACTCAAG 290
Db      841 ACCCAACAGGGAAGAAAGTTAATGCGAAGTGTGCAATGCTTATCCAGAGAACTCAAG 900
QY      291 CTTGGAAGCGAGCTCTCCAGGAGCTATTGCAACCTTTGAAGCAGAGTTGCGTTTACAG 350
Db      901 CTTGGAAGCGAGCTCTCCAGGAGCTATTGCAACCTTTGAAGCAGAGTTGCGTTTACAG 960
QY      351 AGAAATACAGTGAAGAGCTTTAAAGCAGTCAGATGAATGAATGCTTCAATCATCCAG 410
Db      961 AGAAATACAGTGAAGAGCTTTAAAGCAGTCAGATGAATGAATGCTTCAATCATCCAG 1020
QY      411 CTTGATGAAGAAGTAGAGGTATGCAAGTAGTACCATTCTAGTTCTGCGAGCAGCTGAAG 470
Db      1021 CTTGATGAAGAAGTAGAGGTATGCAAGTAGTACCATTCTAGTTCTGCGAGCAGCTGAAG 1080
QY      471 GAGACACCCAGCAGTTGGCTCAGTACCAGCAGCAGCTCTCAGGCTCTGCCCCAAAGT 530
Db      1081 GAGACACCCAGCAGTTGGCTCAGTACCAGCAGCAGCTCTCAGGCTCTGCCCCAAAGT 1140
QY      531 ACCAGCAGGACTACAGCTTCTGAACTGTAGAACAGTCAGAGGCCACAAAGTAAAGACTGC 590
Db      1141 ACCAGCAGGACTACAGCTTCTGAACTGTAGAACAGTCAGAGGCCACAAAGTAAAGACTGC 1200
QY      591 AGTCGTCTGACAAACGGACCAAGTAAATGGTAGCTCTCTCCGCCAGAGGAGCTCTGGGTCT 650
Db      1201 AGTCGTCTGACAAACGGACCAAGTAAATGGTAGCTCTCTCCGCCAGAGGAGCTCTGGGTCT 1260
QY      651 GGATTTTCACAGGAGGGGCAACCAACCGAAGATGATTTCTCTCTCTCCAGGGAATGGT 710
Db      1261 GGATTTTCACAGGAGGGGCAACCAACCGAAGATGATTTCTCTCTCTCCAGGGAATGGT 1320
QY      711 AATAAGTCTTCAACAGCTCAGAGGAGAGAACTGGCAGAGGAGGTAGTGGTTACGTAAAT 770
Db      1321 AATAAGTCTTCAACAGCTCAGAGGAGAGAACTGGCAGAGGAGGTAGTGGTTACGTAAAT 1380
QY      771 CAATCAGTGGGGGTATGAAGTGTAGACTCTCCACCGGGCAGTGAAGAACTCTCTCACA 830
Db      1381 CAATCAGTGGGGGTATGAAGTGTAGACTCTCCACCGGGCAGTGAAGAACTCTCTCACA 1440
QY      831 CACCAATCAAAATGACACAGACTCCAGTCCATGACCTCAAGAGGAGAAAGCAGTGGG 890
Db      1441 CACCAATCAAAATGACACAGACTCCAGTCCATGACCTCAAGAGGAGAAAGCAGTGGG 1500
QY      891 AAAGTAATCGAATGCTGGGTTCGGGCCAGCTTCCAGAGTTCAGAGTGGTGAAGTGAAT 950

```

Db 1501 AAGAGTAATCGAACTGTGGGTTCCCGCACGTTTCCAGATGGCTTGGACTCAAGTGAAT 1560

Qy 951 GTACAGGGTTCAGTTTGTAAATATTTTTCAGCAAAATTTTATACAGTGTCAATTAATTT 1010

Db 1561 GTACAGGGTTCAGTTTGTAAATATTTTTCAGCAAAATTTTATACAGTGTCAATTAATTT 1620

Qy 1011 GGGAGGATACGTCCAGAAATTAATGCATCTTTTGTCAAAATTCGCTTTTGTGG 1070

Db 1621 GGGAGGATACGTCCAGAAATTAATGCATCTTTTGTCAAAATTCGCTTTTGTGG 1680

Qy 1071 GTGTACGTTTGGTTTTTTTGTGTTTTTTTGTGTTTTTTTGTGTTTTTTTGTGTTTTTTT 1130

Db 1681 GTGTACGTTTGGTTTTTTTGTGTTTTTTTGTGTTTTTTTGTGTTTTTTTGTGTTTTTTT 1740

Qy 1131 TAAGAGTGTTGTGTTTTTTTGTGTTTTTTTGTGTTTTTTTGTGTTTTTTTGTGTTTTTTT 1190

Db 1741 T---TTTTTTTTTTTTTTTTTGTGTTTTTTTGTGTTTTTTTGTGTTTTTTTGTGTTTTTTT 1797

Qy 1191 TAATTACTTTGAATGTTC-AAAAGGCAATTTTGTGAGGGTCCAGTTA 1238

Db 1798 TAATTACTTTGAATGTTC-AAAAGGCAATTTTGTGAGGGTCCAGTTA 1846

RESULT 7

HSA276706 1167 bp mRNA linear PRI 26-SEP-2000

LOCUS Homo sapiens partial mRNA for WTAP protein.

DEFINITION HSA276706

ACCESSION AJ276706

VERSION AJ276706.1 GI:10334525

KEYWORDS Wtap gene; WTAP protein.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 Little, N.A. Thesis (1999) Edinburgh University

REFERENCE 2 Little, N.A., Hastie, N.D. and Davies, R.C. Identification of WTAP, a novel Wilms' tumour 1-associated protein Hum. Mol. Genet. 9 (15), 2231-2239 (2000)

JOURNAL Hum. Mol. Genet. 9 (15), 2231-2239 (2000)

MEDLINE 20458888

PUBMED 11001926

REFERENCE 3 (bases 1 to 1167)

AUTHORS Hastie, N.D.

TITLE Direct Submission

JOURNAL Submitted (22-MAR-2000) Hastie N.D., Developmental Biology, MRC Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh, EH4 2XU, UNITED KINGDOM

FEATURES

source

1..1167

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone_lib="Clontech human foetal kidney MATCHMAKER cDNA library"

gene

CDS

1..1167

/gene="wtap"

<1..>1167

/gene="wtap"

/codon_start=2

/product="WTAP protein"

/protein_id="CAC10188.1"

/db_xref="GI:10334525"

/db_xref="TrEMBL:Q96T28"

/translation="EPELPKVRISLSTDFKVMARDELILRWQVEAYVQALEGYTDL NSNDVTGRSESEELKQOQESARRENILVMRLATKEQWQECTTQYLKQVQPSV AOLSTWYDPAINLPFLKMGLEQTKDLFEAONELSAWKFDPDSOTGKLMKACRM LIQNEQLGRQLSGRIALQAEALOKYSEELKSSQDELNDPFIOLDREVEGMOST ILVLCQQLKTRQQLAQKQOQESASPSISRTTASBPVQSEATSKDCSLRTGPSN GSSSRQRTSGSGFHREGNTTDDFPSPGNGKNSNSSEERTGGRGGYVNLQSLAGYE SVDEPTGSGNSLTQSDNTDSDHPQBEKAVSGKGNRTVGRHVQNGILDSSNV"

ORIGIN

RESULT 8

AL135914

LOCUS

AL135914

165802 bp

DNA

linear

PRI 20-JUL-2001

Query Match 76.2%; Score 943.8; DB 9; Length 1167;

Best Local Similarity 99.3%; Pred. No. 9.2e-204; Indels 0; Gaps 0;

Matches 948; Conservative 0; Mismatches 7;

Qy 1 ACATCCCTTGGAAATCGCACTAGCAACCAAGGACAAAGATATTCAAGAGTGTACTACTCAAA 60

Db 213 ACATCCCTTGGAAATCGCACTAGCAACCAAGGACAAAGATATTCAAGAGTGTACTACTCAAA 272

Qy 61 TCCAGTACTCAAGCAAGTCCAGCAGCTAGCGTTGCCCACTTAGATCAACAAATGGTAG 120

Db 273 TCCAGTACTCAAGCAAGTCCAGCAGCTAGCGTTGCCCACTTAGATCAACAAATGGTAG 332

Qy 121 ACCCAGCCATCAACTTGTTCCTTAAATAATGAAGGTGAATCTGGAACAGACTAAAGACA 180

Db 333 ACCCAGCCATCAACTTGTTCCTTAAATAATGAAGGTGAATCTGGAACAGACTAAAGACA 392

Qy 181 AACTGGAAACAAGCCCAAAATGAATGAGTGCCTGGAAGTTTACGCCTGATAGCCAAACAG 240

Db 393 AACTGGAAACAAGCCCAAAATGAATGAGTGCCTGGAAGTTTACGCCTGATAGCCAAACAG 452

Qy 241 GGAAAAGTTAATGCGGAAGTGTGAATGCTTATCCAGGAGAAATCAAGAGCTTGGAAAGC 300

Db 453 GGAAAAGTTAATGCGGAAGTGTGAATGCTTATCCAGGAGAAATCAAGAGCTTGGAAAGC 512

Qy 301 AGCTCTCCAGGGACGTAATGCACAACCTTGAAGCAGAGTTGGCTTTACAGAGAAATACA 360

Db 513 AGCTCTCCAGGGACGTAATGCACAACCTTGAAGCAGAGTTGGCTTTACAGAGAAATACA 572

Qy 361 GTGAGAGCTTAAAGCAGTCCAGGATGAATGACTTTCATCATCAGCTTGTATGAAG 420

Db 573 GTGAGAGCTTAAAGCAGTCCAGGATGAATGACTTTCATCATCAGCTTGTATGAAG 632

Qy 421 AAGTAGAGGGTATGACAGAGTACCAATTCCTTCTGACAGCAGCTGAAGAGACACGCC 480

Db 633 AAGTAGAGGGTATGACAGAGTACCAATTCCTTCTGACAGCAGCTGAAGAGACACGCC 692

Qy 481 AGCAGTTGGCTCAGTACACAGCAGCAGCTCTCAGGCCTCTGCCCAAGTACCACAGCA 540

Db 693 AGCAGTTGGCTCAGTACACAGCAGCAGCTCTCAGGCCTCTGCCCAAGTACCACAGCA 752

Qy 541 CTACAGCTTCTGAACTGTAGAACTGACAGAGGCCACAAAGTAAAGACTGCAGTCTCTGA 600

Db 753 CTACAGCTTCTGAACTGTAGAACTGACAGAGGCCACAAAGTAAAGACTGCAGTCTCTGA 812

Qy 601 CAAACGACCTCAGAGGAGAACTGGCAGAGGAGTAGTGGTTACGTAATCAACTCAGTG 660

Db 813 CAAACGACCTCAGAGGAGAACTGGCAGAGGAGTAGTGGTTACGTAATCAACTCAGTG 872

Qy 661 GGGAGGGCAACACACCGAAGATGACTTCTCTTCTCCAGGGAATGGTAAATAGTCTT 720

Db 873 GGGAGGGCAACACACCGAAGATGACTTCTCTTCTCTCCAGGGAATGGTAAATAGTCTT 932

Qy 721 CCAACAGCTCAGAGGAGAACTGGCAGAGGAGTAGTGGTTACGTAATCAACTCAGTG 780

Db 933 CCAACAGCTCAGAGGAGAACTGGCAGAGGAGTAGTGGTTACGTAATCAACTCAGTG 992

Qy 781 CGGGGTATCAAAAGTGTAGACTCTCCACGGGCAAGTGAATCTCTCTCACACACCAATCAA 840

Db 993 CGGGGTATCAAAAGTGTAGACTCTCCACGGGCAAGTGAATCTCTCTCACACACCAATCAA 1052

Qy 841 ATGACACAGACTCCAGTCAATGACCTTCAAGAGGAGAAAGCAGTGTGGGAAAGTAAATC 900

Db 1053 ATGACACAGACTCCAGTCAATGACCTTCAAGAGGAGAAAGCAGTGTGGGAAAGTAAATC 1112

Qy 901 GAACTGTGGGTTCCCGCCACGTTTCAAGATGGCTTGGACTCAAGTGTAAATGTACA 955

Db 1113 GAACTGTGGGTTCCCGCCACGTTTCAAGATGGCTTGGACTCAAGTGTAAATGTACA 1167


```

DEFINITION      Human DNA sequence from clone RPl-56L9 on chromosome 6, complete
sequence.
ACCESSION      AL135914
VERSION        AL135914.25  GI:15020718
KEYWORDS       HTG.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 165802)
Williams, S.
Direct Submission
Submitted (20_JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerques@sanger.ac.uk
On Jul 25, 2001 this sequence version replaced gi:14970777.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em:, EMBL; Sw:,
SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
RPl-56L9 is from the library RPC1-1 constructed by the group of
Piet de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2
IMPORTANT: This sequence is not the entire insert of clone RPl-56L9
It may be shorter because we sequence overlapping sections only
once, except for a 100 base overlap.
The true left end of clone RPl-56L9 is at 1 in this sequence. The
true left end of clone RPl-195P10 is at 163803 in this sequence.
The true right end of clone RPl-280I21 is at 123148 in this
sequence.

FEATURES             Location/Qualifiers
     source           1..165802
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RPl-56L9"
                     /clone_lib="RPC1-1"
                     399..564
                     /note="MIR repeat: matches 92..262 of consensus"
                     714..1033
                     /note="AluX repeat: matches 1..303 of consensus"
                     1293..1732
                     /note="L1A16 repeat: matches 5711..6157 of consensus"
                     1755..2035
                     /note="AluSx repeat: matches 23..306 of consensus"
                     2036..2143
                     /note="54 copies 2 mer ga 72% conserved"
                     2146..2466
                     /note="AluSx repeat: matches 1..310 of consensus"
                     2563..2860
                     /note="MUTLJ repeat: matches 413..513 of consensus"
                     4175..4463
                     /note="AluSg repeat: matches 1..290 of consensus"

```


repeat_region	25044..25357	/note="LIMC4 repeat: matches 6605..6822 of consensus"
repeat_region	25371..26043	/note="LIMC4 repeat: matches 7167..7977 of consensus"
repeat_region	26147..26273	/note="FLAM A repeat: matches 19..142 of consensus"
repeat_region	26898..27558	/note="LIPBA repeat: matches -1271..-531 of consensus"
repeat_region	27661..30938	/note="LIPB1 repeat: matches -279..2897 of consensus"
repeat_region	30939..31236	/note="ALUSg repeat: matches 1..296 of consensus"
repeat_region	31237..32134	/note="LIPB1 repeat: matches 2897..3766 of consensus"
repeat_region	32135..32930	/note="SVA repeat: matches 1..791 of consensus"
repeat_region	32931..33685	/note="LIPB1 repeat: matches 3766..5563 of consensus"
repeat_region	33694..34341	/note="LTR12 repeat: matches 17..670 of consensus"
repeat_region	34342..34477	/note="LIPB1 repeat: matches 5559..5681 of consensus"
repeat_region	34478..34782	/note="ALUSg repeat: matches 1..306 of consensus"
repeat_region	34783..34889	/note="LIPB1 repeat: matches 5681..5791 of consensus"
repeat_region	34892..34927	/note="Alu repeat: matches 1..36 of consensus"
repeat_region	34931..35025	/note="LIPB1 repeat: matches 5687..5791 of consensus"
repeat_region	35062..35339	/note="139 copies 2 mer at 63% conserved"
repeat_region	35340..35699	/note="LIPB1 repeat: matches 5781..6152 of consensus"
repeat_region	35762..35807	/note="23 copies 2 mer ct 91% conserved"
repeat_region	35850..35973	/note="MLT1A2 repeat: matches 243..374 of consensus"
repeat_region	36221..36687	/note="L2 repeat: matches 2165..2726 of consensus"
repeat_region	36812..36888	/note="L2 repeat: matches 1910..1982 of consensus"
repeat_region	36889..37175	/note="ALUSp repeat: matches 1..296 of consensus"
repeat_region	37176..37379	/note="L2 repeat: matches 1689..1910 of consensus"
repeat_region	38041..38195	/note="AluJb repeat: matches 140..294 of consensus"
repeat_region	38304..38462	/note="Aluuo repeat: matches 1..134 of consensus"
repeat_region	38859..38896	/note="MERSB repeat: matches 4..35 of consensus"
repeat_region	38897..39182	
Query March 64.4%; Score 797.8; DB 9; Length 165802;		
Best Local Similarity 97.5%; Pred. No. 1.6e-170;		
Matches 832; Conservative 0; Mismatches 17; Indels 4; Gaps 2;		
Qy	387	GAACCTGAATGACTTCATCCAGCTTGATGAAGAAGTAGAGGGTATCCAGAGTACCATT 446
Db	126710	GAACCTGAATGACTTCATCCAGCTTGATGAAGAAGTAGAGGGTATCCAGAGTACCATT 126769
Qy	447	CTAGTTCTGCAGCAGCACTGAAGAGACAGCCAGCAGTTGGCTCAGTACCAGCAGAG 506
Db	126770	CTAGTTCTGCAGCAGCACTGAAGAGACAGCCAGCAGTTGGCTCAGTACCAGCAGAG 126829
Qy	507	CAGTCTCAGGCTCTGCCCCAAGTACCAGCAGACTACAGCTTCTGAACCTGTAGAACAG 566
Db	126830	CAGTCTCAGGCTCTGCCCCAAGTACCAGCAGACTACAGCTTCTGAACCTGTAGAACAG 126889
Qy	567	TCAGAGGCCAACAGTAAGACTGTGAGTCTGTCTGAACAACGACCAAGTAATCGTAGTCC 626
Db	126890	TCAGAGGCCAACAGTAAGACTGTGAGTCTGTCTGAACAACGACCAAGTAATCGTAGTCC 126949

The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
 4 (bases 1 to 203077)
 Wileg,G., Yang,L., Prescott,A., Qin,B. and Roe,B.A.
 Direct Submission
 Submitted (06-DEC-2002) Department Of Chemistry And Biochemistry,
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
 OK 73019, USA
 5 (bases 1 to 203077)
 Wileg,G., Yang,L., Prescott,A., Qin,B., Brathwaite,M., Waelz,P.,
 Nagaraja,R. and Roe,B.A.
 Direct Submission
 Submitted (21-MAR-2003) Department Of Chemistry And Biochemistry,
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
 OK 73019, USA
 6 (bases 1 to 203077)
 Wileg,G., Yang,L., Prescott,A., Qin,B., Brathwaite,M., Waelz,P.,
 Nagaraja,R. and Roe,B.A.
 Direct Submission
 Submitted (18-APR-2003) Department Of Chemistry And Biochemistry,
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
 OK 73019, USA
 7 (bases 1 to 203077)
 Wileg,G., Yang,L., Prescott,A., Qin,B., Brathwaite,M., Waelz,P.,
 Nagaraja,R. and Roe,B.A.
 Direct Submission
 Submitted (16-SEP-2003) Department Of Chemistry And Biochemistry,
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
 OK 73019, USA
 On Apr 18, 2003 this sequence version replaced gi:26080527.

 Center: Department Of Chemistry And Biochemistry
 The University Of Oklahoma
 Center code:UOKNOR

Location/Qualifiers
 1. 203077
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /map="17"
 /clone="rp23-387d2"
 /clone_lib="RPCI - 23 Female (C57BL/6J) Mouse BAC Library"

Query Match 53.8%; Score 666.6; DB 10; Length 203077;
 Best Local Similarity 89.8%; Pred. No. 1e-140;
 Matches 766; Conservative 0; Mismatches 69; Indels 18; Gaps 4;

Qy 387 GAACGATGACCTTCATCCAGCTTGATGAAGAGTACGAGGTATGAGAGTACCAT 446
 Db 135004 GAACGATGACCTTCATCCAGCTTGATGAAGAGTACGAGGTATGAGAGTACCAT 134945
 Qy 447 CTAGTTCTGAGCAGCAGCTGAAGAGACGCGCAGCTTGCTCAGTACCGAGCAG 506
 Db 134944 CTAGTTCTTTCAGCAACAATTGAAGAGACGCGCAGCTTGCTCAGTACCGAGCAG 134825
 Qy 507 CAGTCTCAGGCTCTGCCCCAAGTACCAGCAGCTACAGCTTCTCAACTGTAGACAG 566
 Db 134884 CAGTCTCAGGCTCAGCTCCAGTACAGCAGCTACATCTTCTGAACCTGTAGATCAG 134825
 Qy 567 TCAGAGCCCAAGTAAGACTCGAGTCTGCTGCAACAGGACCAAGTAATGTAGCTCC 626
 Db 134824 GCAGAGGTCAAGCAAGACTCGAGTCTGCTGCAACAGGACCAAGTAATGTAGCTCC 134765
 Qy 627 TCCGCCAGAGGAGCTGCTGGTCTGGATTTCAGAGGAGGCGACACACCGAGATGAC 686
 Db 134764 TCCGCCAGAGGAGCTGCTGGTCTGGATTTCAGAGGAGGCGAGCACCTGAGGATGAC 134705
 Qy 687 TTTCCTTCTTCTCCAGGGAATGGTAATAAGTCTCCCAACAGCTTCAGAGGAGAGAACTGGC 746
 Db 134704 TTTCCTTCTTCTCCAGGGAATGGTAATAAGGCTCCCAACAGCTTCAGAGGAGAGAACTGGC 134645

Qy 747 AGAGGAGGTAGTGGTTACGTAATCAACTCAGTGGGGGTATGAAAGTGTAGACTCTCC 806
 Db 134644 AGAGGAGGTAGTGGTTACGTAATCAACTCAGTGGGGGTATGAAAGTGTAGACTCTCC 134585
 Qy 807 ACGGCGAGTGAAAACCTCTCTCACACCAATCAAAATGACACAGCTCCAGTCTACACCT 866
 Db 134584 ACGGCGAGTGAAAACCTCTCTCACACCAATCAAAATGACACAGCTCCAGTCTACACCT 134525
 Qy 867 CAAGAGGAGGAGGAGTGGAGTGGGAAAGTAATCGAACTGGGGTCCGCGCACCTTCAG 926
 Db 134524 CAAGAGGAGGAGGAGTGGAGTGGGAAAGTAATCGAACTGGGGTCCGCGCATTTTCAG 134465
 Qy 927 AATGGCTTGGACTCAAGTGTAAATGTACAGGGTTCAGTTTTGTATATATTTTTCAGCAA 986
 Db 134464 AATGGCTTGGACTCAAGTGTAAATGTACAGGGTTCAGTTTTGTATATATTTTTCAGCAA 134405
 Qy 987 TTTTATACAGTGTCAATTAATTTGGGAGAGATCTCTCCAGAAAATAATGCAATCT 1046
 Db 134404 TTTTATACAGTGTCAATTAATTTGGGAGAGATCTCTCTCCAGAAAATAATGCAATCT 134346
 Qy 1047 TTGTACCAATTTGCCCTTTTGTGGGTGACGTTTTGTGGTTTTTTTGTGTTTTTTTCT 1106
 Db 134345 TTGTACCAATTTTACCTTTTGTGGGTGACGTTTTGTGGTTTTTTTGTGTTTTTTTCT 134288
 Qy 1107 TTGGAATTTTTTCTTTTCTTTTAAAGAGGTGGTGTGTTTTTTTGTCTCAATCTTC 1166
 Db 134287 TTTTATT 134242
 Qy 1167 TCGCGGCTGGGAATTTGAACAGTAAATTAATTAATTTGAATGTTCC-AAAAGGCATTTGTG 1225
 Db 134241 TCGCACCTTGAAATTTGAACAGTAAATTAATTTGAACATTTGCTAAAAGGACATTTGTG 134182
 Qy 1226 GAGGCTCCAGTTA 1238
 Db 134181 TAGGGTCAAGTTA 134169

RESULT 10
 AL589878
 LOCUS
 DEFINITION Mouse DNA sequence from clone RP23-351C7 on chromosome 15, complete sequence.
 ACCESSION AL589878
 VERSION AL589878.12 GI:20068450
 KEYWORDS HTG.
 SOURCE Mus musculus (house mouse)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 Phillimore,B.
 Direct Submission
 Submitted (04-APR-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 On Apr 7, 2002 this sequence version replaced gi:16304732.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP

database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-351C7 is
from the RPI-23 Mouse PAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6.

FEATURES

Location/Qualifiers
1..203949
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="15"
/clone="RP23-351C7"
/clone_lib="RPI-23"

ORIGIN

Query Match 53.8%; Score 666.6; DB 10; Length 203949;
Best Local Similarity 89.8%; Pred. No. 1e-140;
Matches 766; Conservative 0; Mismatches 69; Indels 18; Gaps 4;
387 GAACCTGAATGACCTTCATCATCCAGCTTGATGAAGAAGTAGAGGGTATGCAGAGTACCATT 446
Db 68594 GAACCTGAATGACCTTCATCATCCAGCTTGATGAAGAAGTAGAGGGTATGCAGAGTACCATT 68653
387 GAACCTGAATGACCTTCATCATCCAGCTTGATGAAGAAGTAGAGGGTATGCAGAGTACCATT 446
Db 68594 GAACCTGAATGACCTTCATCATCCAGCTTGATGAAGAAGTAGAGGGTATGCAGAGTACCATT 68653
447 CTAGTTCTGCAGCAGCAGCTGAAGGAGACACGCCAGCAGTTGGCTCAGTACCGAGCAGCAG 506
Db 68654 CTAGTTCTGCAGCAGCAGCTGAAGGAGACACGCCAGCAGTTGGCTCAGTACCGAGCAGCAG 506
507 CAGTCTCAGCGCTCTGCCCAAGTACCAGCAGGACTACAGCTTCTGAACCTGTGAGAACAG 566
Db 68714 CAGTCTCAGCGCTCTGCCCAAGTACCAGCAGGACTACAGCTTCTGAACCTGTGAGAACAG 566
567 TCAGAGGCCACAACTGAAGACTGCAGTCGTCTGACAAACCGACCAAGTAATGTAGTCTCC 626
Db 68774 GCAGAGGTCAACAGCAAGACTGCAGTCGTCTGACAAACCGACCAAGTAATGTAGTCTCC 68833
627 TCCCGCCAGAGCAGCTCTGGTCTGGATTTACAGAGGAGGCGACACAAACCGAGAGTAC 686
Db 68834 TCCCGCCAGAGCAGCTCTGGTCTGGATTTACAGAGGAGGCGACACAAACCGAGAGTAC 68893
687 TTTCCTTCTTCCAGGGAATGGTAAATAGTCTCTCCAAACAGCTCAGAGGAGAGAACTGGC 746
Db 68894 TTTCCTTCTTCCAGGGAATGGTAAATAGTCTCTCCAAACAGCTCAGAGGAGAGAACTGGC 68953
747 AGAGAGGTAGTGGTTAGTAAATCACTCAGTCCGGGGTATGAAGTGTAGTCTCC 806
Db 68954 AGAGAGGTAGTGGTTAGTAAATCACTCAGTCCGGGGTATGAAGTGTAGTCTCC 69013
807 ACGGCACTGAAACTCTCTCACACCAATCAATGACACAGACTCCAGTCAATGACCCCT 866
Db 69014 ACGGCACTGAAACTCTCTCACACCAATCAATGACACAGACTCCAGTCAATGACCCCT 69073
867 CAAGAGGAGAAAGCAGTGGGAAAGGTAAATCGAACTGGGTTCGCCGACGTTTCAG 926
Db 69074 CAAGAGGAGAAAGCAGTGGGAAAGGTAAATCGAACTGGGTTCGCCGACGTTTCAG 69133
927 AATGCTTGGACTCAAGTGAATAGTACAGGTTTCAGTTTGTAAATATTTTTCAGCAA 986
Db 69134 AATGCTTGGACTCAAGTGAATAGTACAGGTTTCAGTTTGTAAATATTTTTCAGCAA 69193
987 TTTTATACAGTGTCAATTAATTTGGGAGAGATCTGTCCAGAAATTAATGATACATT 1046
Db 69194 TTTTATACAGTGTCAATTAATTTGGGAGAGATCTGTCCAGAAATTAATGATACATT 69252
1047 TTGTCACAATTTGCCCTTTTGGGGTGTACGTTTGGTTTGTGTTTGTGTTTGTGTTTCT 1106
Db 69253 TTGTCACAATTTGCCCTTTTGGGGTGTACGTTTGGTTTGTGTTTGTGTTTGTGTTTCT 69310
1107 TTGGATTTTCTTTCTTTTAAAGAGGTGGTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTCT 1166
Db 69311 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 108685
1167 TCGCGCTGGGAATGTGAACAGTAAATTAATTTGAAATGTTGC-AAAAGGCAATTTGTG 1225

69357 TGCCACTTTGAAATTTGTAACAGTTAATTAATTTGAAACATTTGCTAAAGGACATTTTGTG 69416
QY 1226 GAGGGTCCAGTTA 1238
Db 69417 TAGGGTCAAGTTA 69429
RESULT 11
AC074222.1/c
WPCOMMENT
Sequence split into 5 fragments LOCUS AC074222 Accession AC074222
Fragment Name Begin End
AC074222.0 1 110000
AC074222.1 100001 210000
AC074222.2 200001 310000
AC074222.3 300001 410000
AC074222.4 400001 412547
Continuation (2 of 5) of AC074222 from base 100001 (AC074222 Mus musculus chromosome 17
Query Match 52.9%; Score 654.6; DB 2; Length 110000;
Best Local Similarity 89.7%; Pred. No. 5.4e-138;
Matches 765; Conservative 0; Mismatches 69; Indels 19; Gaps 5;
387 GAACCTGAATGACCTTCATCATCCAGCTTGATGAAGAAGTAGAGGGTATGCAGAGTACCATT 446
Db 109400 GAACCTGAATGACCTTCATCATCCAGCTTGATGAAGAAGTAGAGGGTATGCAGAGTACCATT 109341
447 CTAGTTCTGCAGCAGCAGCTGAAGGAGACACGCCAGCAGTTGGCTCAGTACCGAGCAGCAG 506
Db 109340 CTAG-TCTTCAGCAAAATTTGAAGGAGACACGCCAGCAGTTGGCAGCAGTACCGAGCAG 109282
507 CAGTCTCAGCGCTCTGCCCAAGTACCAGCAGGACTACAGCTTCTGAACCTGTGAGAACAG 566
Db 109281 CAGTCTCAGCGCTCTGCCCAAGTACCAGCAGGACTACAGCTTCTGAACCTGTGAGAACAG 109222
567 TCAGAGGCCACAACTGAAGACTGCAGTCGTCTGACAAACCGACCAAGTAATGTAGTCTCC 626
Db 109221 GCAGAGGTCAACAGCAAGACTGCAGTCGTCTGACAAACCGACCAAGTAATGTAGTCTCC 109162
627 TCCCGCCAGAGCAGCTCTGGTCTGGATTTACAGAGGAGGCGACACAAACCGAGAGTAC 686
Db 109161 TCCCGCCAGAGCAGCTCTGGTCTGGATTTACAGAGGAGGCGACACAACTGAGAGTAC 109102
687 TTTCCTTCTTCCAGGGAATGGTAAATAGTCTCTCCAAACAGCTCAGAGGAGAGAACTGGC 746
Db 109101 TTTCCTTCTTCCAGGGAATGGTAAATAGTCTCTCCAAACAGCTCAGAGGAGAGAACTGGC 109042
747 AGAGAGGTAGTGGTTAGTAAATCACTCAGTCCGGGGTATGAAGTGTAGTCTCC 806
Db 109041 AGAGAGGTAGTGGTTAGTAAATCACTCAGTCCGGGGTATGAAGTGTAGTCTCC 108982
807 ACGGCACTGAAACTCTCTCACACCAATCAATGACACAGACTCCAGTCAATGACCCCT 866
Db 108981 ACGGCACTGAAACTCTCTCACACCAATCAATGACACAGACTCCAGTCAATGACCCCT 108922
867 CAAGAGGAGAAAGCAGTGGGAAAGGTAAATCGAACTGGGTTCGCCGACGTTTCAG 926
Db 108921 CAAGAGGAGAAAGCAGTGGGAAAGGTAAATCGAACTGGGTTCGCCGACGTTTCAG 108862
927 AATGCTTGGACTCAAGTGAATAGTACAGGTTTCAGTTTGTAAATATTTTTCAGCAA 986
Db 108861 AATGCTTGGACTCAAGTGAATAGTACAGGTTTCAGTTTGTAAATATTTTTCAGCAA 108802
987 TTTTATACAGTGTCAATTAATTTGGGAGAGATCTGTCCAGAAATTAATGATACATT 1046
Db 108801 TTTTATACAGTGTCAATTAATTTGGGAGAGATCTGTCCAGAAATTAATGATACATT 108743
1047 TTGTCACAATTTGCCCTTTTGGGGTGTACGTTTGGTTTGTGTTTGTGTTTGTGTTTCT 1106
Db 108742 TTGTCACAATTTGCCCTTTTGGGGTGTACGTTTGGTTTGTGTTTGTGTTTGTGTTTCT 108685
1107 TTGATTTTCTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 1166

[illegible]

```

||||| 685 CTTGATCAGGAGGTAGAGGGTATGCAGATACCAATCTAGTCTTCAGCAGAGTTGAG 744
||||| QY 471 GAGACAGCGCAGAGTTGGCTAGTACCAGCAGCAGCAGTCTCAGGCGCTCTGCCCAAGT 530
||||| Db 745 GAGACTCGCAGCAGTTGGCGAGTACAGCAGCAGCAGTCCAGGCGCTCCACCCAGT 804
||||| QY 531 ACCAGCAGGACTACAGTCTTGACCTGTAGAACAGTACAGGCGCACAAGTAAAGACTGC 590
||||| Db 805 ACCAGCAGGACTCCATCTTCTGAGCCTACAGACAGGAGAGGCTGTGGGTAAAGACTGC 864
||||| QY 591 AGTCGCTGACAAACGAGCAAGTAATGTTAGTCTCTCCGCGCAGAGGAGCTGTGGGTCT 650
||||| Db 865 AGCGCGCTGCTAAGCAGCCACAGATGTTAGTCTCTCCATCAGCGGAGCTGTGGGTCT 924
||||| QY 651 GGATTTTCACAGGAGGCGCAACACCAAGAGATGATTTCTCTTCTTCAGGGAATGTT 710
||||| Db 925 GGATTTTATAGGAGGAGGTAGCAGACAGAGATGATCTCCGGCTTCTCCAGGGAATGTT 984
||||| QY 711 AATAAGTCTCCACAGCTCAGAGGAGAGAACTGGCAGAGGAGTGTAGTACGTAAAT 770
||||| Db 985 AATAAGCTGTCCAAACCACTCTGAAGATAGAACTGGTAGAGAGTGTGTAGTACATAAC 1044
||||| QY 771 CAACTCAGTGGGGGTATGAAGTGTAGACTCTCCACGGGCGAGTGAAGAACTCTTCACA 830
||||| Db 1045 CAACTCAGTACTGGGTATGAAGTGTAGACTCTCCACCTGCGAGTGAAGAACTCTCTCACT 1104
||||| QY 831 CACCAATCAATGACACAGACTCAGTATGACCTCAGCAGGAGAGAAAGCAGTGTGGG 890
||||| Db 1105 CACCATTCAATGACACAGACTCCTCAATCATGATCTCCTCAAGAGGAGAGAAAGCAGTGTGGG 1164
||||| QY 891 AAGGTAATCGAAGTGTGGGTTCCGCGCAGCTTCCAGAACTGGTGGACTCAAGTGTAAAT 950
||||| Db 1165 AAGGTAACAGAACTGGGTTCTCGCATGTCCAGAACTGGTGGACTCAAGTGTAAAT 1224
||||| QY 951 GTACAGGGTTCAGTTTGTAAATTTTTCAGCAAAATTTTATACAGTGTCAATTTT 1010
||||| Db 1225 GTGAGGGTTCAGTTTGTAAATTTTTCGAGCAATTTTATACAGTGTCAATTTT 1284
||||| QY 1011 GGGAGAGTACTCTCCAGA-----AAATTAATGATATCTTTTGCACAAATTCG 1060
||||| Db 1285 GGGAGAGTACTCTCTCAGAAAGCGTTATTAATTAATGATGATCTTGTCAATTTTGC 1344
||||| QY 1061 CTTTTTGTGGTGTACGTTT 1080
||||| Db 1345 CTTTTTGTGGATTTCTTTT 1364

```

```

RESULT 14
AC112423
LOCUS Rattus norvegicus clone CH230-29021, WORKING DRAFT SEQUENCE, 7
DEFINITION AC112423 165096 bp DNA linear HTG 15-NOV-2002
unordered pieces.
ACCESSION AC112423 GI:25006897
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 165096)
Munry,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alébrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,

```

Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Bean, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, C., Guerra, W., Guevara, M., Guevarne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Loresuhewa, L., Loulseghe, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwackelmeah, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Popper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.L., Puzo, M., Quiroz, J., Rachin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, R., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Wang, V., Villanueva, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, R., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Center: Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GRST
Center clone name: CH230-29021

TITLE
JOURNAL
REFERENCE
2 (bases 1 to 165096)
Unpublished
Worley, K.C.
Direct Submission
Submitted (21-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 165096)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Nov 15, 2002 this sequence version replaced gi:23270045.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

COMMENT

```

----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 143564 bases at least Q40
Consensus quality: 146570 bases at least Q30
Consensus quality: 148264 bases at least Q20
Estimated insert size: 138236; sum-of-contigs estimation
Quality coverage: 9x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
  (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).
* NOTE: This is a 'working draft' sequence. It currently
  consists of 7 contigs. The true order of the pieces
  is not known and their order in this sequence record is
  arbitrary. Gaps between the contigs are represented as
  runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
  as soon as it is available and the accession number will
  be preserved.
*
1
3091: contig of 3091 bp in length
3092
3191: gap of unknown length
3192
31741: contig of 28550 bp in length
31742
31841: gap of unknown length
31842
97626: contig of 65785 bp in length
97627
97726: gap of unknown length
97727
102107: contig of 4381 bp in length
102108
102207: gap of unknown length
102208
162803: contig of 60596 bp in length
162804
162903: gap of unknown length
162904
162986: contig of 1083 bp in length
162987
164086: gap of unknown length
164087
165096: contig of 1010 bp in length.
164087
165096: contig of 1010 bp in length.
-----
FEATURES             Location/Qualifiers
     1..165096
         /organism="Rattus norvegicus"
         /mol_type="genomic DNA"
         /db_xref="taxon:10116"
         /clone="CH230-29021"
     misc_feature      31842..33215
         /note="wgs_contig"
     misc_feature      121972..123198
         /note="wgs_contig"
-----
ORIGIN
Query Match          49.2%; Score 608.6; DB 2; Length 165096;
Best Local Similarity 85.3%; Pred No 1.ee-127;
Matches 736; Conservative 0; Mismatches 74; Indels 43; Gaps 4;
QY 387 GAACGTAATGCTTCATCATCCAGCTTGATGAAGAAGTAGAGGGTATGCAGAGTACCAT 446
DB 46563 GAACGTAATGCTTCATCATCCAGCTTGATGAAGAAGTAGAGGGTATGCAGAGCACCATT 46622
QY 447 CTAGTCTGCAGAGCAGCTGAGGAGACGCCAGCTTGCTCAGTACCGAGCAGAG 506
DB 46623 CTAGTCTTCAGCAACAGTTGAGAGGAGCTCGACAGCGTTGGCAGTACCGAGCAGAG 46682
QY 507 CAGTCTCAGGCCTCTGCCCAAGTACCAGCAGAGCTACAGCTTCTGAACTGTAGAACAG 566
DB 46683 CAGTCTCAGGCTTCTGCTCCAGTACCAGCAGGACTACATCTTCTGAACTGTGATCAG 46742
QY 567 TCAGAGGCCACAGTAAGAGCTCAGTCTGCTGCAAAAGGACCAAGTAAATGGTAGCTCC 626
DB 46743 GCAGAGGCCACAGCAAGAGCTCAGTCTGCTGCAAAAGGACCAAGTAAATGGTAGCTCC 46802
QY 627 TCCCGCCAGAGCAGCTGTGGGTCTGGATTTCACAGGAGGCGCAACACACCGAAGATGAC 686
DB 46803 TCCCGCCAGAGCAGCTGTGGGTCTGGATTTCACAGGAGGCGGAGCAGCTCTCTGAGATGAC 46862
QY 687 TTTCCTTTCTCCAGGAAATGTAATAGTCTCTCAAGAGGAGGAGAGAACTGGC 746
DB 46863 TTTCCTTTCTCTTCAGGGAACGGTAATAAGGCGCTCCACAGCTCAGAGGAGAACTGGC 46922
QY 747 AGAGGAGGTAGTGGTTACGTAAATCAACTCAGTCAGTCGGGGTATGAAGTGTAGACTCTCC 806
DB

```

Lorensuhera, L., Louleaged, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartine, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munitasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokeleneh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Popper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shety, J., Shvartbeyn, A., Sisson, I., Sitter, C. D., Smales, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Zhen, D., von Niederhausen, A., Weis, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G., and Gibbs, R. A.

Direct Submission
 Unpublished
 2 (bases 1 to 246585)
 Worley, K.C.

Direct Submission
 Submitted (03-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 246585)
 Rat Genome Sequencing Consortium.

Direct Submission
 Submitted (09-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On May 9, 2003 this sequence version replaced gi:24818213.
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GCEM
 Center clone name: CH230-6511
 ----- Summary Statistics
 Assembly program: Atlas
 Consensus quality: 229387 bases at least Q40
 Consensus quality: 232288 bases at least Q30
 Consensus quality: 234515 bases at least Q20
 Estimated insert size: 23986; sum-of-contigs estimation
 Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 3 contigs. The true order of the pieces
 * is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 236514: contig of 236514 bp in length
 * 236515 236614: gap of unknown length
 * 236615 245383: contig of 8769 bp in length
 * 245384 245384: gap of unknown length
 * 245484 245585: contig of 1102 bp in length.
 Location/Qualifiers
 1. 245585
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-6E11"
 1. 1085
 /note="wgs end extension
 clone_end:Sp6"
 1816..2431
 /note="clone boundary
 clone_end:Sp6
 site:ECORI
 end sequence: BH358751"
 complement(229899..230437)
 /note="clone boundary
 clone_end:T7
 site:ECORI
 end sequence: BH358748"
 236615..239647
 /note="wgs end extension
 clone_end:T7"
 243668..245383
 /note="wgs end extension
 clone_end:T7"

ORIGIN
 Query Match 49.2%; Score 608.6; DB 2; Length 246585;
 Best Local Similarity 86.3%; Pred. No. 1.6e-127;
 Matches 736; Conservative 0; Mismatches 74; Indels 43; Gaps 4;
 QY 387 GAACGTAATGATTCATATCCAGCTTGAAGAGAGTACAGGAGTACGAGTACCAT 446
 Db 193449 GAACGTAATGATTCATATCCAGCTTGAAGAGAGTACAGGAGTACGAGTACCAT 193508
 QY 447 CTAGTCTCAGCAGCAGCTGAAGAGAGACACCCAGCAGTTCGCTCAGTACCAAGCAGCAG 506
 Db 193509 CTAGTCTCAGCAGCAGCTGAAGAGAGTACAGCAGTTCGCTCAGTACCAAGCAGCAG 193568
 QY 507 CAGTCTCAGGCGCTCTGCCCAAGTACCAAGTACCAAGTACCAAGTACCAAGTACCAAG 566
 Db 193569 CAGTCTCAGGCGCTCTGCCCAAGTACCAAGTACCAAGTACCAAGTACCAAGTACCAAG 193628
 QY 567 TCGAGGCCACAAAGTAAAGCTGCTGCTGCAAAAGGACCAAGTAAAGTAAAGTAAAGT 626
 Db 193629 TCGAGGCCACAAAGTAAAGCTGCTGCTGCAAAAGGACCAAGTAAAGTAAAGTAAAGT 193688
 QY 627 TCCCGCCAGAGGACGCTCTGGGCTCTGGATTTTCAGAGGGGGGCAACACACCAAGATGAC 686
 Db 193689 TCCCGCCAGAGGACGCTCTGGGCTCTGGATTTTCAGAGGGGGGAGGACCTCTCAGGATGAC 193748
 QY 687 TTTCCTTTCTTCCAGGGAATGTAATAGTCTTCAAGCAGCTCAGAGGAGAGACTGGC 746
 Db 193749 TTTCCTTTCTTCCAGGGAATGTAATAGTCTTCAAGCAGCTCAGAGGAGAGACTGGC 193808
 QY 747 AGAGGAGGAGTGGTTTACGTAAATCAACTCAGTGGGGGTATGAAAGTGTAGACTCTCC 806
 Db 193809 A----GAGGTAGTGTAGTACATAAACCCACTCAGTGGGGGTATGAAAGTGTAGACTCTCC 193865
 QY 807 ACGGGCAGTGAAGTCTCTTCACACACCAATCAATGACAGACTCCAGTCTGACCT 866
 Db 193866 ACGGGCAGTGAAGTCTCTCTCACACATCACTCAATGACAGACTCCAGTCTGACCT 193925

```
QY      867  CAAGAGGAGAAAGCAGTGTGGGAAAGGTAATCGAACTGTGGGTCCCGCCACGTTCCAG  926
Db      193926 CAAGAGGAGAAAGCAGTGTGGGAAAGGTAATCGAACTGTGGGTCCCGCCACGTTCCAG  193985
QY      927  AATGGCTTGGACTCAAGTGTAAATGTACAGGGTTCAGTTTTGTAAATTTTTTCAGCAAA  986
Db      193986 AATGGCTTGGACTCAAGTGTAAATGTACAGGGTTCAGTTTTGTAAATTTTTTCAGCAAA  194045
QY      987  TTTTATACAGTGTCAATTAATTTGGGAGAGGATCTGTCCAGAAAATTAATGCATCTT  1046
Db      194046 TTTTATACAGTGTCAATTAATTTGGGAGAGGATCTGTCCAGAAAATTAATGCATCTT  194103
QY      1047 TTGTCACAATTTGCCCTTTTGTGGGTGTACGTTTTGTTTTTTTTTTTGTGTTTTTTTCT  1106
Db      194104 TTGTCACAATTTATACCCCTTTTGTGGGTGTGTTTTGTTTTGTTGTTCTTTTTTTT---  194160
QY      1107 TTGGAATTTTTTCTTTCTTTTAAAGAGGTGTTGTGTTTTTGTCTTCAATCTTC  1166
Db      194161 -----TTTTTTTAAATGCTTCAATACCTC  194186
QY      1167 TCGCGCTTGGAAATGTAAAGATTACTTTGAATGTGC-AAAAGGCATTTTGTG  1225
Db      194187 TGCTACTTTGAAAATGTAAAGATTACTTTGAACATTTGCTAAAAGGACATTTTGTG  194246
QY      1226 CAGGTCAGTTA  1238
Db      194247 TAGGTCAGTTA  194259
```

Search completed: November 14, 2004, 12:54:06
Job time : 5531 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2004, 11:21:58 ; Search time 1973 Seconds
(without alignments)
3388.962 Million cell updates/sec

Title: US-10-030-389-12

Perfect score: 1238

Sequence: 1 acatcttggaatgcgacta.....ttttgtgagggtccagta 1238

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3625171 seqs, 2700493622 residues

Total number of hits satisfying chosen parameters: 7250342

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
- 18: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 19: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
- 20: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 21: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1116.6	90.2	2772	15	US-10-252-157-398
2	797.4	64.4	41309	17	US-10-741-601-5771
3	379.8	30.7	31657	17	US-10-741-601-5690
4	329.8	26.6	12648	17	US-10-741-601-5652
5	296.6	24.0	440	9	US-09-983-965-217
6	226.2	18.3	480	16	US-10-242-535A-15856
7	226.2	18.3	480	16	US-10-085-783A-15856
8	226.2	18.3	1600	9	US-09-840-787-50
9	226.2	18.3	1600	13	US-10-044-090-550
10	226.2	18.3	1622	9	US-09-954-456-492
11	226.2	18.3	1622	15	US-10-172-118-1100
12	226.2	18.3	1622	16	US-10-342-887-1100
					Sequence 398, App
					Sequence 5771, App
					Sequence 5690, App
					Sequence 5652, App
					Sequence 217, App
					Sequence 15856, A
					Sequence 15856, A
					Sequence 50, Appl
					Sequence 550, App
					Sequence 492, App
					Sequence 1100, App
					Sequence 1100, App

13	226.2	18.3	1691	13	US-10-044-090-552	Sequence 552, App
14	226.2	18.3	1996	13	US-10-044-090-551	Sequence 551, App
15	201.2	16.8	694	18	US-10-425-115-6885	Sequence 6885, App
16	201.2	16.3	475	9	US-09-867-701-2417	Sequence 2417, App
17	200.6	16.2	201	17	US-10-741-601-25687	Sequence 25687, A
18	191.4	15.5	601	14	US-10-066-543-1379	Sequence 1379, App
19	150.6	12.2	201	17	US-10-741-601-25688	Sequence 25688, A
20	143.4	11.6	3902	13	US-10-044-090-451	Sequence 451, App
21	143.4	11.6	3902	15	US-10-084-817-260	Sequence 260, App
22	134.8	10.9	201	17	US-10-741-601-14585	Sequence 14585, A
23	134.8	10.9	201	17	US-10-741-601-25705	Sequence 25705, A
24	120.8	9.8	2642	16	US-10-108-260A-2162	Sequence 2162, App
25	120.8	9.8	81826	14	US-10-175-523-197	Sequence 197, App
26	115.2	9.3	501	9	US-09-833-790-179	Sequence 179, App
27	108	8.7	481	10	US-09-918-395-29597	Sequence 29597, A
28	93	7.5	558	9	US-09-728-446-797	Sequence 797, App
29	75	6.1	501	9	US-09-833-790-168	Sequence 168, App
30	75	6.1	33053	17	US-10-433-793-35	Sequence 35, Appl
31	72.4	5.8	6794	16	US-10-613-133	Sequence 133, App
32	71.4	5.8	40324	17	US-10-433-793-179	Sequence 179, App
33	69	5.6	5823	15	US-10-340-453-256	Sequence 256, App
34	68.5	5.5	785	17	US-10-767-701-8312	Sequence 8312, App
35	68	5.5	12007	15	US-10-311-455-690	Sequence 690, App
36	67	5.4	201	17	US-10-741-601-25715	Sequence 25715, A
37	67	5.4	18624	15	US-10-311-455-1676	Sequence 1676, App
38	66.3	5.4	5413	16	US-10-221-714A-418	Sequence 418, App
39	66.5	5.4	411	9	US-09-960-352-14521	Sequence 14521, A
40	66.5	5.4	447	18	US-10-425-115-148961	Sequence 148961, A
41	66.5	5.4	6668	15	US-10-311-455-1669	Sequence 1669, App
42	66.5	5.4	3673778	15	US-10-312-841-1	Sequence 1, Appli
43	66.4	5.4	516	17	US-10-021-323-13996	Sequence 13996, A
44	66.2	5.3	554	18	US-10-425-115-139665	Sequence 139665, A
45	66.2	5.3	1287	16	US-10-424-599-97419	Sequence 97419, A

ALIGNMENTS

RESULT 1

US-10-252-157-398
Sequence 398, Application US/10252157
Publication No. US20030190640A1
GENERAL INFORMATION:
APPLICANT: Paris, Mary
TITLE OF INVENTION: GENES EXPRESSED IN PROSTATE CANCER
FILE REFERENCE: PA-0027-1 US
CURRENT APPLICATION NUMBER: US/10/252,157
PRIOR FILING DATE: 2002-10-01
PRIOR APPLICATION NUMBER: 60/295,048
PRIOR FILING DATE: 2001-05-31
NUMBER OF SEQ ID NOS: 501
SOFTWARE: PERL Program
SEQ ID NO 398
LENGTH: 2772
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20030190640A1 040057.2
US-10-252-157-398

Query Match	90.2%	Score	1116.6	DB	15	Length	2772	
Best Local Similarity	99.2%	Pred. No.	3.2e-265					
Matches 1122	Conservative	0	Mismatches	9	Indels	0	Gaps	0
QY	1	ACATCTTGGAAATGCGACTAGCAACCAAGGCAAGATATTCAGAGTGTACTACTCAA	60					
Db	1641	ACATCTTGGTAATGCGACTAGCAACCAAGGCAAGATATTCAGAGTGTACTACTCAA	1700					
QY	61	TCCAGTACTCAAGCAAGTCAGAGCCTAGCGTGTGCCCACTTAGATCAACAATGCTAG	120					
Db	1701	TCCAGTACTCAAGCAAGTCAGAGCCTAGCGTGTGCCCACTTAGATCAACAATGCTAG	1760					

```
QY 121 ACCAGCGATCAACTGTGTTTTTCCCTAAATAAGTGAAGTGAACAGACTAAAGACA 180
Db 1761 ACCAGCGATCAACTGTGTTTTTCCCTAAATAAGTGAAGTGAACAGACTAAAGACA 1820
QY 181 AACTGGAACAGCCCAAAATGAAGTGTGCTGGAAGTTTACGCTGATAGCCAAACAG 240
Db 1821 AACTGGAACAGCCCAAAATGAAGTGTGCTGGAAGTTTACGCTGATAGCCAAACAG 1880
QY 241 GGAATAAGTAAATGCGAAGTGTGCAATGCTTATCCAGGAGAATCAAGAGCTTGAAGGC 300
Db 1881 GGAATAAGTAAATGCGAAGTGTGCAATGCTTATCCAGGAGAATCAAGAGCTTGAAGGC 1940
QY 301 AGCTGTCAGGAGGAGTATGACAACTGGAAGAGTGTGCTTACAGAGAAATACA 360
Db 1941 AGCTGTCAGGAGGAGTATGACAACTGGAAGAGTGTGCTTACAGAGAAATACA 2000
QY 361 GTGAGGAGCTTAAAGACAGTCAAGTGAATGACTTCAATCATCCAGCTTCAATGAAG 420
Db 2001 GTGAGGAGCTTAAAGACAGTCAAGTGAATGACTTCAATCATCCAGCTTCAATGAAG 2060
QY 421 AAGTAGAGGATGACAGAGTACCATTCTAGTTCTGACAGCAGCTGAAGAGACAGCC 480
Db 2061 AAGTAGAGGATGACAGAGTACCATTCTAGTTCTGACAGCAGCTGAAGAGACAGCC 2120
QY 481 AGCAGTGGCTCAGTACCAGCAGCAGTCTCAGGCTCTGCCCAAGTACCAGCAGGA 540
Db 2121 AGCAGTGGCTCAGTACCAGCAGCAGTCTCAGGCTCTGCCCAAGTACCAGCAGGA 2180
QY 541 CTACAGCTTCTGAACCTGTGAAACAGTCAAGGAGCCCAAGTAAAGACTGCGTCTGA 600
Db 2181 CTACAGCTTCTGAACCTGTGAAACAGTCAAGGAGCCCAAGTAAAGACTGCGTCTGA 2240
QY 601 CAACCGGACCAAGTAAAGTGTAGTCTCCCGCAGAGGAGTCTGGGTCTGGATTTACA 660
Db 2241 CAACCGGACCAAGTAAAGTGTAGTCTCCCGCAGAGGAGTCTGGGTCTGGATTTACA 2300
QY 661 GGGAGGGCAACACAAACGAAAGTGAATGCTTCTCCAGGGAATGGTAAAGTCT 720
Db 2301 GGGAGGGCAACACAAACGAAAGTGAATGCTTCTCCAGGGAATGGTAAAGTCT 2360
QY 721 CCACAGCTCAGAGGAGAGTCTCCACCGGAGTGAATGCTTCTCCAGGGAATGGTAAAGTCT 780
Db 2361 CCACAGCTCAGAGGAGAGTCTCCACCGGAGTGAATGCTTCTCCAGGGAATGGTAAAGTCT 2420
QY 781 CGGGGTATGAAGTGTAGACTCTCCACCGGAGTGAATGCTTCTCCACCGGAGTGAATGCT 840
Db 2421 CGGGGTATGAAGTGTAGACTCTCCACCGGAGTGAATGCTTCTCCACCGGAGTGAATGCT 2480
QY 841 ATGACAGAGTCCAGTCAATGACCTCAAGAGGAGAAAGCAGTGAAGTGAAGTAAATC 900
Db 2481 ATGACAGAGTCCAGTCAATGACCTCAAGAGGAGAAAGCAGTGAAGTGAAGTAAATC 2540
QY 901 GAAGTGTGGGTTCGCGCCAGTTCAGAAATGGCTTGAAGTGAATGCTTCAAGTGAAGT 960
Db 2541 GAAGTGTGGGTTCGCGCCAGTTCAGAAATGGCTTGAAGTGAATGCTTCAAGTGAAGT 2600
QY 961 CAGTTTGTAAATTTTTCAGCAAAATTTTATACAGTGTCAATTTAAATTTGGAGAGAT 1020
Db 2601 CAGTTTGTAAATTTTTCAGCAAAATTTTATACAGTGTCAATTTAAATTTGGAGAGAT 2660
QY 1021 ACTGTCCAGAAATTAATGATCACTTTTGTCAAAATTTTGTGGGTGTAGCTTT 1080
Db 2661 ACTGTCCAGAAATTAATGATCACTTTTGTCAAAATTTTGTGGGTGTAGCTTT 2720
QY 1081 TGGTTTTTTTGTGTTTTTTTCTTGGATTTTTTTCTTTCTTTTCTTTT 1131
Db 2721 TGGTTTTTTTGTGTTTTTTTCTTGGATTTTTTTCTTTCTTTTCTTTT 2771
```

RESULT 2

US-10-741-601-5771

: Sequence 5771, Application US/10741601

```
: Publication No. US20040166519A1
: GENERAL INFORMATION:
: APPLICANT: CARBELL, Michele et al.
: TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
: TITLE OF INVENTION: STENOSIS, METHODS OF DETECTION AND USES THEREOF
: FILE REFERENCE: CL001500
: CURRENT APPLICATION NUMBER: US/10/741,601
: CURRENT FILING DATE: 2003-12-22
: NUMBER OF SEQ ID NOS: 26415
: SOFTWARE: Fast-Seq for Windows Version 4.0
: SEQ ID NO 5771
: LENGTH: 41309
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-10-741-601-5771
```

```
Query Match 64.4%; Score 797.4; DB 17; Length 41309;
Best Local Similarity 97.4%; Pred. No. 185;
Matches 831; Conservative 1; Mismatches 17; Indels 4; Gaps 2;

QY 387 GAACCTGAATGACTTCATCCAGCTTGTGAAGAAGTAGAGGTATGACAGTACCACTT 446
Db 34018 GAACCTGAATGACTTCATCCAGCTTGTGAAGAAGTAGAGGTATGACAGTACCACTT 34077
QY 447 CTAGTTCTGCAGCAGCAGCTGAAAGAGACACGCCAGCAGTTGGCTCAGTACAGCAG 506
Db 34078 CTAGTTCTGCAGCAGCAGCTGAAAGAGACACGCCAGCAGTTGGCTCAGTACAGCAG 34137
QY 507 CAGTCTCAGGCTCTGCCCTCCAGTACCAGCAGGACTCAGCTTCTGAACCTGTAGAAC 566
Db 34138 CAGTCTCAGGCTCTGCCCTCCAGTACCAGCAGGACTCAGCTTCTGAACCTGTAGAAC 34197
QY 567 TCAGAGGCCACAAAGTAAAGACTGCTGTCTGAACCGGACCAAGTAATGTTAGTCT 626
Db 34198 TCAGAGGCCACAAAGTAAAGACTGCTGTCTGAACCGGACCAAGTAATGTTAGTCT 34257
QY 627 TCCGCCCAGAGAGCTCTGGGTCTGGATTTCACAGGGAGGGCAACACACCGAGAGTAC 686
Db 34258 TCCGCCCAGAGAGCTCTGGGTCTGGATTTCACAGGGAGGGCAACACACCGAGAGTAC 34317
QY 687 TTTCTCTCTTCTCCAGGGAATGGTAAATGAAGTCTTCAACAGCTCAGAGGAGAGTGGC 746
Db 34318 TTTCTCTCTTCTCCAGGGAATGGTAAATGAAGTCTTCAACAGCTCAGAGGAGAGTGGC 34377
QY 747 AGAGAGGTAGTGTGTAATCAACTCAGTCTCGGGGTATGAAAGTGTAGACTCTCCC 806
Db 34378 AGAGAGGTAGTGTGTAATCAACTCAGTCTCGGGGTATGAAAGTGTAGACTCTCCC 34437
QY 807 ACGGCAGTGAATACTCTCTCACACCAATCAATGACAGACTCCAGTCAATGACCTT 866
Db 34438 ACGGCAGTGAATACTCTCTCACACCAATCAATGACAGACTCCAGTCAATGACCTT 34497
QY 867 CAAGAGGAGAAAGCAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 926
Db 34498 CAAGAGGAGAAAGCAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 34557
QY 927 AATGGCTTGGACTCAAGTGAATAATGTACAGGTTTCAAGTTTGTAAATTTTTCAGAAA 986
Db 34558 AATGGCTTGGACTCAAGTGAATAATGTACAGGTTTCAAGTTTGTAAATTTTTCAGAAA 34617
QY 987 TTTTATACAGTGTCAATTAATTTGGAGAGGATATCTGTCCAGAAATTAATGCACTT 1046
Db 34618 TTTTATACAGTGTCAATTAATTTGGAGAGGATATCTGTCCAGAAATTAATGCACTT 34677
QY 1047 TTGTCAAAATTTGCTTTTGTGGGTGTAGCTTTTGTGTTTTTGTGTTTTTGTGTTTTTCT 1106
Db 34678 TTGTCAAAATTTGCTTTTGTGGGTGTAGCTTTTGTGTTTTTGTGTTTTTGTGTTTTTCT 34737
QY 1107 TTGATTTTTTTTCTTTTCTTTTAAAGAGGTTGTTGTTTTTGTGTTTTTGTGTTCAATCTTC 1166
Db 34738 TTGATTTTTTTTCTTTTCTTTTAAAGAGGTTGTTGTTTTTGTGTTTTTGTGTTCAATCTTC 34794
QY 1167 TGCAGCTTGGGAATTTGAACAGTTAATTTGATGTTGCTG-AAAAGGGCATTTTGTG 1225
```

Db 34795 TCCCGCTTTGGAATTTGAACAGTTAATTACCTTGAATTTCTCTAAAGGACATTTTGTG 34854
QY 1226 GAGGTCACAGTTA 1238
Db 34855 TAGGGTCAAGTTA 34867

RESULT 3

US-10-741-601-5690
; Sequence 5690, Application US/10741601
; Publication No. US20040166519A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE OF INVENTION: STENOSIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001500
; CURRENT APPLICATION NUMBER: US/10/741,601
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 26415
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5690
; LENGTH: 31657
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-741-601-5690

Query Match 30.7%; Score 379.8; DB 17; Length 31657;
Best Local Similarity 95.2%; Pred. No. 1.4e-82;
Matches 414; Conservative 0; Mismatches 17; Indels 4; Gaps 2;

QY 805 CCACGGGCGAGTGAAGTCTCTACACACCAATCAATGACACAGATCCAGTCATGACC 864
Db 1 CCACGGGCGAGTGAAGTCTCTACACACCAATCAATGACACAGATCCAGTCATGACC 60
QY 865 CTCAGAGGAGAAAGCAGTGCAGTGGGAAAGGTAAATCGAACTGTGGGTCCCGCCACGTTT 924
Db 61 CTCAGAGGAGAAAGCAGTGCAGTGGGAAAGGTAAATCGAACTGTGGGTCCCGCCACGTTT 120
QY 925 AGAATGCTTGGACTCAAGTGTAAATGTAACAGGTTTCAAGTTTGTAAATTTTTCAGCA 384
Db 121 AGAATGCTTGGACTCAAGTGTAAATGTAACAGGTTTCAAGTTTGTAAATTTTTCAGCA 180
QY 985 AATTTTATACAGTGCATTTAATTTGGGAGAGGATCTCCAGAGAAATTAATGATAC 1044
Db 181 AATTTTATACAGTGCATTTAATTTGGGAGAGGATCTCCAGAGAAATTAATGATAC 240
QY 1045 TTTTGTCAATTTGCCCTTTTGTGGGTGACGTTTGGTTTGTGGTTTGTGGTTTGT 1104
Db 241 TTTTGTCAATTTGCCCTTTTGTGGGTGACGTTTGGTTTGTGGTTTGTGGTTTGT 300
QY 1105 CTTTGGATTTTTCCTTTTCTTTTAAAGAGGTTGTTTGTGGTTTGTGGTTTGTGGTTTGT 1164
Db 301 CTTTGGATTTTTCCTTTTCTTTTAAAGAGGTTGTTTGTGGTTTGTGGTTTGTGGTTTGT 357
QY 1165 TCTGGCGCTTGGAAATTTGAACAGTTAATTTACCTTGAATTTTGC-AAAAAGGCAATTTG 1223
Db 358 TCTGGCGCTTGGAAATTTGAACAGTTAATTTACCTTGAATTTTGC-AAAAAGGCAATTTG 417
QY 1224 TGGAGGGTCCAGTTA 1238
Db 418 TGTAGGGTCAAGTTA 432

RESULT 4

US-10-741-601-5652/c
; Sequence 5652, Application US/10741601
; Publication No. US20040166519A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE OF INVENTION: STENOSIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001500

; CURRENT APPLICATION NUMBER: US/10/741,601
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 26415
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5652
; LENGTH: 12648
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-741-601-5652

Query Match 26.6%; Score 329.8; DB 17; Length 12648;
Best Local Similarity 94.5%; Pred. No. 2e-70;
Matches 364; Conservative 0; Mismatches 17; Indels 4; Gaps 2;

QY 855 AGTCATCACCTCAAGAGGAAAGCAGTGCAGTGGGAAAGGTAATCGAACTGTGGGTTC 914
Db 12648 AGTCATCACCTCAAGAGGAAAGCAGTGCAGTGGGAAAGGTAATCGAACTGTGGGTTC 12589
QY 915 CGCCACGTTTCAAGATGGCTGGACTCAAGTGTAAATGACAGGTTTCAGTTTGTAAATAT 974
Db 12588 CGCCACGTTTCAAGATGGCTGGACTCAAGTGTAAATGACAGGTTTCAGTTTGTAAATAT 12529
QY 975 TTTTTCAGCAAAATTTTATACAGTGTCTAATTTAATTTGGGAGAGGATCTGTCAGAAAT 1034
Db 12528 TTTTTCAGCAAAATTTTATACAGTGTCTAATTTAATTTGGGAGAGGATCTGTCAGAAAT 12469
QY 1035 TAATGCATCTTTTGTGCACAAATTTGCCCTTTTGTGGGTGTACGTTTGTGGTTTGT 1094
Db 12468 TAATGCATCTTTTGTGCACAAATTTGCCCTTTTGTGGGTGTACGTTTGTGGTTTGT 12409
QY 1095 TGTCTTTTCTTTTGGAAATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1154
Db 12408 TGTCTTTTCTTTTGGAAATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 12352
QY 1155 CTTCAATCTCTCGCGCTTGGGAAATTTGAACAGTTAATTTACCTTGAATTTGCT-AAAA 1213
Db 12351 CTTCAATCTCTCGCGCTTGGGAAATTTGAACAGTTAATTTACCTTGAATTTGCT-AAAA 12292
QY 1214 GGGCAATTTTGTGGAGGTTCCAGTTA 1238
Db 12291 GGGCAATTTTGTGGAGGTTCCAGTTA 12267

RESULT 5

US-09-983-965-217
; Sequence 217, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Rao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 217
; LENGTH: 440
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 26-BOVMS1-017-Q1-E1-G5
US-09-983-965-217

Query Match 24.0%; Score 296.6; DB 9; Length 440;
Best Local Similarity 89.5%; Pred. No. 5.9e-63;
Matches 332; Conservative 0; Mismatches 34; Indels 5; Gaps 1;

QY 742 CTGGCAGAGGAGGTAGTGGTTACGTAATCAACTCAGTGGCGGTATGAAAGTGTAGACT 801
 Db 1 CTGGCAGAGGAGGTAGTGGTTACGTAATCAACTCAGTGGCGGTATGAAAGTGTAGACT 60
 QY 802 CTCACCGGCGAGTGAATCACTCTCTCACACCAATCAATGACACAGACTCCAGTCATG 861
 Db 61 CTCACCGGCGAGTGAATCACTCTCTCACACCAATCAATGACACAGACTCCAGTCATG 120
 QY 862 ACCCTCAAGAGCAGAGCAGTGGTGAAGTAACTGAGTGGTTCGCGCCACG 921
 Db 121 ACCCGCAGAGGAGAACTCTGAGTGGAAAGTAACTGAGTGGTTCGCGCCACG 180
 QY 922 TTCAGATGGTGGACTCAAGTGTAAATGACAGGGTTCAGTTTGT-----AATATTT 976
 Db 181 TTCAGATGGTGGACTCAAGTGTAAATGACAGGGTTCAGTTTGTAAAAAATAAT 240
 QY 977 TTCAGCAAAATTTTATACAGTGTCAATTAATTTGGAGAGGATAGTTCAGAAATTA 1036
 Db 241 TTCAGCAAAATTTTATACAGTGTCAATTAATTTGGAGAGGATAGTTCAGAAATTA 300
 QY 1037 ATGCATACCTTTGTCACAAATTTGCTTTTGTGGGTGTAGCTTTTGTGTTTTTTTGTG 1096
 Db 301 ATGCATACCTTTGTCACAAATTTGCTTTTGTGGGTGTAGCTTTTGTGTTTTTTTGTG 360
 QY 1097 TTTTTCCTTT 1107
 Db 361 TTTCTTTGCTT 371

RESULT 6
 US-10-242-535A-15856
 ; Sequence 15856, Application US/10242535A
 ; Publication No. US20040013663A1
 ; GENERAL INFORMATION:
 ; APPLICANT: ChondroGene Inc.
 ; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
 ; FILE REFERENCE: 4231/2005
 ; CURRENT APPLICATION NUMBER: US/10/242,535A
 ; CURRENT FILING DATE: 2002-09-12
 ; PRIOR APPLICATION NUMBER: US 10/085,783
 ; PRIOR FILING DATE: 2002-02-28
 ; PRIOR APPLICATION NUMBER: US 60/305,340
 ; PRIOR FILING DATE: 2001-07-13
 ; PRIOR APPLICATION NUMBER: US 60/275,017
 ; PRIOR FILING DATE: 2001-03-12
 ; PRIOR APPLICATION NUMBER: US 60/271,955
 ; PRIOR FILING DATE: 2001-02-28
 ; NUMBER OF SEQ ID NOS: 58994
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 15856
 ; LENGTH: 480
 ; TYPE: DNA
 ; ORGANISM: Human
 US-10-242-535A-15856

Query Match 18.3%; Score 226.2; DB 16; Length 480;
 Best Local Similarity 96.7%; Pred. No. 1.5e-45;
 Matches 231; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
 QY 1 ACATCTCTGGAATGCGACTAGCAACCAAGGACAAAGATATCAAGAGTGTACTACTCAA 60
 Db 51 ACATCTCTGTAATGCGACTAGCAACCAAGGACAAAGATATCAAGAGTGTACTACTCAA 110
 QY 61 TCAGTACTCTCAAGCAAGTCCAGCCGCGAGCGTTCGCCAATGAGATCAACAAATGGTAG 120
 Db 111 TCAGTACTCTCAAGCAAGTCCAGCCGCGAGCGTTCGCCAATGAGATCAACAAATGGTAG 170
 QY 121 ACCAGCGATCAACTGTTTCTTAAATGAAGGTGAAGTGAAGCAAGACTAAAGACA 180
 Db 171 ACCAGCGATCAACTGTTTCTTAAATGAAGGTGAAGTGAAGCAAGACTAAAGACA 230

QY 181 AACTGGAACAAGCCCAAAATGAATGAGTGGCTGGAAGTTTACGCTGATAGCCAAACA 239
 Db 231 AACTGGAACAAGCCCAAAATGAATGAGTGGCTGGAAGTTTACGCTGATAGCCAAACA 289
 RESULT 7
 US-10-085-783A-15856
 ; Sequence 15856, Application US/10085783A
 ; Publication No. US20040037841A1
 ; GENERAL INFORMATION:
 ; APPLICANT: ChondroGene Inc.
 ; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
 ; FILE REFERENCE: 4231/2002
 ; CURRENT APPLICATION NUMBER: US/10/085,783A
 ; CURRENT FILING DATE: 2002-02-28
 ; PRIOR APPLICATION NUMBER: US 60/305,340
 ; PRIOR FILING DATE: 2001-07-13
 ; PRIOR APPLICATION NUMBER: US 60/275,017
 ; PRIOR FILING DATE: 2001-03-12
 ; PRIOR APPLICATION NUMBER: US 60/271,955
 ; PRIOR FILING DATE: 2001-02-28
 ; NUMBER OF SEQ ID NOS: 58994
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 15856
 ; LENGTH: 480
 ; TYPE: DNA
 ; ORGANISM: Human
 US-10-085-783A-15856

Query Match 18.3%; Score 226.2; DB 16; Length 480;
 Best Local Similarity 96.7%; Pred. No. 1.5e-45;
 Matches 231; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
 QY 1 ACATCTCTGGAATGCGACTAGCAACCAAGGACAAAGATATTCAGAGTGTACTACTCAA 60
 Db 51 ACATCTCTGTAATGCGACTAGCAACCAAGGACAAAGATATTCAGAGTGTACTACTCAA 110
 QY 61 TCAGTACTCTCAAGCAAGTCCAGCCGCGAGCGTTCGCCAATGAGATCAACAAATGGTAG 120
 Db 111 TCAGTACTCTCAAGCAAGTCCAGCCGCGAGCGTTCGCCAATGAGATCAACAAATGGTAG 170
 QY 121 ACCAGCGATCAACTGTTTCTTAAATGAAGGTGAAGTGAAGCAAGACTAAAGACA 180
 Db 171 ACCAGCGATCAACTGTTTCTTAAATGAAGGTGAAGTGAAGCAAGACTAAAGACA 230
 QY 181 AACTGGAACAAGCCCAAAATGAATGAGTGGCTGGAAGTTTACGCTGATAGCCAAACA 239
 Db 231 AACTGGAACAAGCCCAAAATGAATGAGTGGCTGGAAGTTTACGCTGATAGCCAAACA 289

RESULT 8
 US-09-840-787-50
 ; Sequence 50, Application US/09840787
 ; Patent No. US20020058264A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lal, Preeti
 ; Hillman, Jennifer L.
 ; Bandman, Olga
 ; Shah, Purvi
 ; Au-Young, Janice
 ; Yue, Henry
 ; Guegler, Karl J.
 ; Corley, Neil C.
 ; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
 ; NUMBER OF SEQUENCES: 98
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.
 ; STREET: 3174 Porter Drive
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/840,787
FILING DATE: 23-Apr-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/518,865
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0356 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-845-0555
TELEFAX: 415-845-4166
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 1600 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: U937NOT01
CLONE: 133
SEQUENCE DESCRIPTION: SEQ ID NO: 50 :
US-09-840-787-50

	Query Match	18.3%	Score 226.2;	DB 9;	Length 1600;
	Best Local Similarity	96.7%;	Pred. No. 2.9e-45;		
Matches	Conservative	0;	Mismatches	8;	Indels
	Gaps	0;			
QY	1	ACATCCTTGGAAATCGCATAGCAACCAAGGACAAGAATTTCAGAGTGCTACTACTCAA	60		
Db	311	ACATCCTTTGTATCGCATAGCAACCAAGGAACAAGAGATGC AAGAGTGCTACTACTCAA	370		
QY	61	TCCAGTAGCTCAAGCAAGTCCAGACGCTAGCGTTGCCAACCTTAGATCAACAATGGTAG	120		
Db	371	TCCAGTAGCTCAAGCAAGTCCAGACGCGAGCGTTGCCAACCTAGATCAACAATGGTAG	430		
QY	121	ACCACGCGATCAACTGTGTTTTCTATAAATGAAAAGTGAAC TGGAAACAGACTAAAGACA	180		
Db	431	ACCACGCGATCAACTGTGTTTTCTATAAATGAAAAGTGAAC TGGAAACAGACTAAAGACA	490		
QY	181	AACTGGAAACAAGCCCCAAAATCAACTGAGTGCGCTGGAA GTTACGCCCTGATAGCAAAAA	239		
Db	491	AACCTGGAAACAAGCCCCAAAATCAACTGAGTGCGCTGGAA GTTACGCCCTGATAGTAAACA	549		

```

RESULT 9
US-10-044-090-550
; Sequence 550, Application US/10044090
; Publication No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 550
; LENGTH: 1600
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 000133CB1
US-10-044-090-550

```

; OTHER INFORMATION: INCYC ID NO: USZ002013,00181
US-10-044-090-550

Query Match	18.3%;	Score	226.2;	DB	13;	Length	1600;
Best Local Similarity	96.7%;	Pred:	No.2.9e-45				
Matches	231;	Conservative	0;	Mismatches	8;	Indels	0;
Gaps	0;						

QY	1	ACATCTTGGAAATCGCACTAGCAACCAAGGCAAGATATTCAGAGTGTACTACTCAA	60
Db	311	ACATCTTGTAAATCGCACTAGCAACCAAGGAACAAGATGCAAGAGTGTACTACTCAA	370
QY	61	TCCAGTACCTCAAGCAAGTCCAGACGCTTAGCGTTGCCCACTTAGATCAACATGGTAG	120
Db	371	TCCAGTACCTCAAGCAAGTCCAGACGCCGAGCGTTGCCCACTGAGATCAACATGGTAG	430
QY	121	ACCCAGCGGATCAACTTGTTTTTTCTTAAAAATGAAAGGTGAACGTGGAACAGACTAAAGACA	180
Db	431	ACCCAGCGGATCAACTTGTTTTTTCTTAAAAATGAAAGGTGAACGTGGAACAGACTAAAGACA	490
QY	181	AAC TGGAACAAGGCCCAAAATGAACTCAGTGCCTTGGAAAGTTTACGCCTCATGCCAAACA	239
Db	491	AAC TGGAACAAGGCCCAAAATGAACTCAGTGCCTTGGAAAGTTTACGCCTCATAGGTAACA	549

```

RESULT 10
US-09-954-456-492
, Sequence 492, Application US/09954456
, Patent No. US20020115057A1
, GENERAL INFORMATION:
, APPLICANT: Young, Paul
, TITLE OF INVENTION: Process for Identifying
, TITLE OF INVENTION: Sets
, FILE REFERENCE: 689290-76
, CURRENT APPLICATION NUMBER: US/09/954,456
, CURRENT FILING DATE: 2001-09-18
, PRIOR APPLICATION NUMBER: US/60/233,617
, PRIOR FILING DATE: 2000-09-19
, PRIOR APPLICATION NUMBER: US/60/234,052
, PRIOR FILING DATE: 2000-09-20
, PRIOR APPLICATION NUMBER: US/60/234,923
, PRIOR FILING DATE: 2000-09-25
, PRIOR APPLICATION NUMBER: US/60/235,134
, PRIOR FILING DATE: 2000-09-25
, PRIOR APPLICATION NUMBER: US/60/235,637
, PRIOR FILING DATE: 2000-09-25
, PRIOR APPLICATION NUMBER: US/60/235,638
, PRIOR FILING DATE: 2000-09-26
, PRIOR APPLICATION NUMBER: US/60/235,711
, PRIOR FILING DATE: 2000-09-27
, PRIOR APPLICATION NUMBER: US/60/235,720
, PRIOR FILING DATE: 2000-09-27
, PRIOR APPLICATION NUMBER: US/60/235,840
, PRIOR FILING DATE: 2000-09-27
, PRIOR APPLICATION NUMBER: US/60/235,863
, PRIOR FILING DATE: 2000-09-27
, NUMBER OF SEQ ID NOS: 2276
, SOFTWARE: PatentIn version 3.0
, SEQ ID NO 492
, LENGTH: 1622
, TYPE: DNA
, ORGANISM: Homo sapiens
US-09-954-456-492

```

	Query Match	18.3%	Score 226.2	DB 9	Length 1622
	Best Local Similarity	96.7%	Pred. No. 2.9e-45		
	Matches 231	Conservative 0	Mismatches 8	Indels 0	Gaps 0
Qy	1	ACATCCCTTGGAAATGCGACTAGCAACCAAGGGACAGATATTTCAAGAGTGTACTACTCAA	60		
Db	345	ACATCCCTTGATGCGACTAGCAACCAAGGAACAAAGAGATGCAAGAGTGTACTACTCAA	404		
Qy	61	TCAGTACCTTCAGCAAGTCCAGCAGCCTAGCGTTGCCCACTTAGATCAACAATGGTAG	120		
Db	405	TCCAGTACCTCAAGCAAGTCCAGCAGCGGAGCGTTGCCCACTGAGATCAACAATGGTAG	464		

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2004, 11:19:34 ; Search time 4402 Seconds
(without alignments)
10248.157 Million cell updates/sec

Title: US-10-030-389-12
Perfect score: 1238
Sequence: 1 acatcttggaatgcacta.....ttttgtggagggtccagtta 1238

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: gb_est1.*
2: gb_est2.*
3: gb_hic.*
4: gb_est3.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_gss1.*
9: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1014.4	81.9	1040	3	HSM802933
2	954.2	77.1	3765	3	CR627456 Homo sapi
3	785.6	63.5	1191	9	AY416358 Homo sapi
4	702.4	56.7	1191	9	AY416360 Mus muscu
5	692.8	56.0	860	5	BU138450 603135077
6	622.8	50.3	1130	5	BK395380 BX395380
7	608	49.1	710	7	CN166400 997672 MA
8	604.4	48.8	758	4	B1089654 602854348
9	602.6	48.7	808	7	CO387170 AGNCOURT
10	595.6	48.1	710	5	BU942029 AGNCOURT
11	594.2	48.0	656	6	CB419819 592705 MA
12	590.6	47.7	702	4	B1185059 UNL-P-FN-
13	571.8	46.2	880	5	BU200958 603949240
14	566.4	45.8	660	1	AJ657839 AGNCOURT
15	553.4	44.7	961	6	CB182874 AGNCOURT
16	547.4	44.2	549	2	AW936487 QV4-DT002
17	545.6	44.1	883	5	BU109878 603127357
18	543.6	43.9	925	1	AL895534 AL895534
19	541.2	43.7	835	7	CO561762 AGNCOURT
20	537.8	43.4	757	7	CO389204 AGNCOURT
21	537.6	43.4	680	7	CK308195 SB02045B2
22	536.2	43.3	884	5	BU139100 603134268
23	529.6	42.8	544	2	AW753139 RCI-CT024
24	528.8	42.7	575	2	AW936415 QV4-DT002

25	518	41.8	947	5	BU132234	BU132234	603117225
26	510	41.2	606	4	BG896779	BG896779	HOA50-1-B
27	503.4	41.1	907	7	CF586305	CF586305	AGENCOURT
28	508.6	40.9	612	4	BG924897	BG924897	HNC38-1-E
29	505.8	40.9	862	5	BF165439	BF165439	BP165439
30	496	40.1	840	5	EX771427	EX771427	EX771427
31	492.2	39.8	805	5	BU110001	BU110001	603126168
32	489.8	39.6	617	5	BU143216	BU143216	603228940
33	489.6	39.5	637	1	AJ398971	AJ398971	AJ398971
34	485.4	39.2	700	1	AJ397697	AJ397697	AJ397697
35	477.2	38.5	799	5	BX665454	BX665454	BX665454
36	476	38.4	910	7	CF412751	CF412751	CH38081-G
37	471	38.0	548	6	CB287321	CB287321	CMD57_505
38	471	38.0	788	5	BU244773	BU244773	603779112
39	469	37.9	1161	5	BU137716	BU137716	603125072
40	461.2	37.3	667	6	CF254734	CF254734	mdvnl19-G
41	456.2	36.8	586	2	BF930054	BF930054	MR2-NT013
42	452.2	36.5	582	2	BF930039	BF930039	MR2-NT013
43	450.2	36.4	1031	5	BK387163	BK387163	BK387163
44	441.8	35.7	550	4	BG142863	BG142863	la9e05-Y
45	441.2	35.6	814	6	CF258081	CF258081	phan105_f

ALIGNMENTS

RESULT 1
HSM802933 1040 bp mRNA linear HTC 03-AUG-2004
LOCUS Homo sapiens mRNA; cDNA DKFZp761K0722 (from clone DKFZp761K0722).
DEFINITION
ACCESSION AL583911
VERSION AL583911.1 GI:13093774
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1040)
Wambutt,R., Heubner,D., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
Fobo,G., Han,M. and Wiemann,S.
The German cDNA Consortium
Submitted (03-AUG-2004) MIPS, Ingolstaedter Landstr.1, D-85764
Neuherberg, GERMANY
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by Agowa (Berlin/Germany) within the cDNA sequencing
consortium of the German Genome Project.
This clone (DKFZp761K0722) is available at the RZPD Deutsches
Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
Please contact RZPD for ordering:
http://www.rzpd.de/cgi-bin/products/cl.cgi?cloneID=DKFZp761K0722
Further information about the clone and the sequencing project is
available at http://mips.gsf.de/projects/cdna/.

FEATURES
Location/Qualifiers
1..1040
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="RZPD:DKFZp761K0722"
/db_xref="taxon:9606"
/clone="DKFZp761K0722"
/tissue_type="amygdala"
/clone_lib="761 (synonym: hamy2). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="adult"
/note="Wilms' tumour 1-associating protein isoform 1,
N-terminus truncated"
1..1040
/gene="DKFZp761K0722"
3..860
/genes="DKFZp761K0722"
/codon_start=1
/product="hypothetical protein"

QY 231 AGCCAAACAGGGAAGTTAATGGCGAAGTGTGCAATGCTTATCCAGAGAAATCAAGAG 290
Db 2258 AGCCAAACAGGGAAGTTAATGGCGAAGTGTGCAATGCTTATCCAGAGAAATCAAGAG 2317
QY 291 CTTGGAAGGAGCTGTCCAGGAGCTATTGGCAACTTGAAGCAGAGTTGGCTTTACAG 350
Db 2318 CTTGGAAGGAGCTGTCCAGGAGCTATTGGCAACTTGAAGCAGAGTTGGCTTTACAG 2377
QY 351 AAGAAATACAGTGAAGGAGCTTAAAGAGCTGAGATGAATGAATGACTTCATCCAG 410
Db 2378 AAGAAATACAGTGAAGGAGCTTAAAGAGCTGAGATGAATGAATGACTTCATCCAG 2437
QY 411 CTTGATGAAGAGTGAAGGATGTCAGAGTACCATTTCTAGTTCTGAGCAGAGCTGAAG 470
Db 2438 CTTGATGAAGAGTGAAGGATGTCAGAGTACCATTTCTAGTTCTGAGCAGAGCTGAAG 2497
QY 471 GAGACAGCCAGAGTTGGCTCAGTACCAGCAGCAGCAGTCTCAGGCTCTGCCCCAAGT 530
Db 2498 GAGACAGCCAGAGTTGGCTCAGTACCAGCAGCAGCAGTCTCAGGCTCTGCCCCAAGT 2557
QY 531 ACCAGCAGGACTACAGTCTTGAACCTGTAGACAGTACAGGAGCCAGTAAGACTGC 590
Db 2558 ACCAGCAGGACTACAGTCTTGAACCTGTAGACAGTACAGGAGCCAGTAAGACTGC 2617
QY 591 AGTCGTCTGCAAAACGACCAAGTAATGGTAGCTCTCCGCGCAGAGGAGCTGTGGGTCT 650
Db 2618 AGTCGTCTGCAAAACGACCAAGTAATGGTAGCTCTCCGCGCAGAGGAGCTGTGGGTCT 2677
QY 651 GGATTTACAGGGAGGAGCAACACAAACGAGATGACTTCTCTCTCTCAGGGAATGTT 710
Db 2678 GGATTTACAGGGAGGAGCAACACAAACGAGATGACTTCTCTCTCTCAGGGAATGTT 2737
QY 711 AATAGTCTCCACAGCTCAGAGGAGAGAACTGGCAGAGAGTGTAGTGTACGTAAT 770
Db 2738 AATAGTCTCCACAGCTCAGAGGAGAGAACTGGCAGAGAGTGTAGTGTACGTAAT 2797
QY 771 CAATCAGTGGGGATGAAAGTGTAGACTCTCCACGGGAGTGAAGAACTCTCTACA 830
Db 2798 CAATCAGTGGGGATGAAAGTGTAGACTCTCCACGGGAGTGAAGAACTCTCTACA 2857
QY 831 CACCAATCAATGACAGAGCTCAGTCAATGACCTCAGAGGAGAGAGAGTGTGGGG 890
Db 2858 CACCAATCAATGACAGAGCTCAGTCAATGACCTCAGAGGAGAGAGAGTGTGGGG 2917
QY 891 AAGGTAAATCGAATGTGGGTTCCGCGCAGCTTCCAGAACTGGTGGACTCAAGTAAAT 950
Db 2918 AAGGTAAATCGAATGTGGGTTCCGCGCAGCTTCCAGAACTGGTGGACTCAAGTAAAT 2977
QY 951 GTACAGGTTGAGTTTGTAAATATTTTTCAGGAAATTTTATACAGTGTCAATTAAT 1010
Db 2978 GTACAGGTTGAGTTTGTAAATATTTTTCAGGAAATTTTATACAGTGTCAATTAAT 3037
QY 1011 GGGAGAGGATCTGTCAGAGAAATTAATGATATCTTTTGTACAAATTTGCTTTTGTGG 1070
Db 3038 GGGAGAGGATCTGTCAGAGAAATTAATGATATCTTTTGTACAAATTTGCTTTTGTGG 3097
QY 1071 GTGACGTTTGGTTTTTTTGTGTTTTTTTCTTTGGAATTTTTTTCTTTCTTTT 1130
Db 3098 GTGACGTTTGGTTTTTTTGTGTTTTTTTCTTTGGAATTTTTTTCTTTCTTTT 3157
QY 1131 TAAGAGGTTGTTGTTTTTTTCTTCAATFACCTTCTGCGGCTTGGGAATTTGAACAGT 1190
Db 3158 T- -TTTTTTTTTTTTTTTTTTTGGTTTCAATFACCTTCTGCGGCTTGGGAATTTGAACAGT 3215
QY 1191 TAATTAATTTGAATTTTC- AAAAGGGCAATTTTGGAGGCTCCAGTTA 1238
Db 3216 TAATTAATTTGAATTTTC- AAAAGGGCAATTTTGGAGGCTCCAGTTA 1264

RESULT 3
AY416358
LOCUS

AY416358 1191 bp DNA linear GSS 12-DEC-2003

DEFINITION Homo sapiens HCM5864 gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION AY416358 GI:39772318
VERSION AY416358.1
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1191)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 1191)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source
1..1191
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
<1..>1191
/locus_tag="HCM5864"
ORIGIN
Query Match 63.5%; Score 785.6; DB 9; Length 1191;
Best Local Similarity 81.2%; Pred. No. 5.9e-174;
Matches 783; Conservative 0; Mismatches 183; Indels 0; Gaps 0;
QY 1 ACATCTCTTGGATGGAGCTAGCAACCAAGGAGCAGAGATATTCAAGAGTGTACTCTCAA 60
Db 221 ACATCTCTTGGATGGAGCTAGCAACCAAGGAGCAGAGATATTCAAGAGTGTACTCTCAA 280
QY 61 TCCAGTACCTCAAGCAAGTCCAGAGCTGAGCTGCTGCCAACTTAGATCAACAATGGTAG 120
Db 281 NNN 340
QY 121 ACCCAGCGATCAACTGTTTTCCTTAAATAATGAAAGGTGAAGTGAAGTGAAGTGAAG 180
Db 341 NNN 400
QY 181 AACTGGACCAAGCCCAAAATGAATGAGTGGCTGGAGTTTACGCTGATAGCCAAACAG 240
Db 401 NNN 460
QY 241 GGAAAAAGTTATGGGAGTGTGGAATGCTTATCCAGGAGAAATCAAGAGCTTGAAGGC 300
Db 461 GGAAAAAGTTATGGGAGTGTGGAATGCTTATCCAGGAGAAATCAAGAGCTTGAAGGC 520
QY 301 AGCTGTCCCGAGGAGTATTGCACAACTTTGAAGCAGAGTTGGCTTTACAGAGAAATACA 360
Db 521 AGCTGTCCCGAGGAGTATTGCACAACTTTGAAGCAGAGTTGGCTTTACAGAGAAATACA 580
QY 361 GTGAGAGCTTAAAGCAGTCAAGATGAATGACTTCTATCATCTCCAGCTTGTAGAG 420
Db 581 GTGAGAGCTTAAAGCAGTCAAGATGAATGACTTCTATCATCTCCAGCTTGTAGAG 640
QY 421 AAGTAGAGGTATGCAAGAGTACCAATTTCTAGTTCTGCGAGCAGAGTGTGAAGGAGACGCG 480
Db 641 AAGTAGAGGTATGCAAGAGTACCAATTTCTAGTTCTGCGAGCAGAGTGTGAAGGAGACGCG 700


```

Db      1181  CAGTTTGTAA 1191
|||||
RESULT 5
BUI38450      860 bp  mRNA  linear  EST 25-NOV-2002
LOCUS        603135077F1 CSBQCHL24 Gallus gallus cDNA clone ChEST117j19 5', mRNA
DEFINITION
ACCESSION    BUI38450
VERSION      BUI38450.1 GI:25352616
KEYWORDS     EST.
SOURCE       Gallus gallus (chicken)
ORGANISM     Gallus gallus
REFERENCE    1 (bases 1 to 860)
AUTHORS      Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
              Fong,W.I., Fickler,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
TITLE        A Comprehensive Collection of Chicken cDNAs
JOURNAL      Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE      22335534
PUBMED       12445392
COMMENT      Contact: Simon Hubbard
              Department of Biomolecular Sciences
              University of Manchester Institute of Science and Technology
              (UMIST)
              PO Box 88, Manchester, M60 1QD, UK
              Tel: 01612008930
              Fax: 01612360409
              Email: Simon.Hubbard@umist.ac.uk.
FEATURES     source
              Location/Qualifiers
              1..860
                /organism="Gallus gallus"
                /mol_type="mRNA"
                /strain="White Leghorn, Hisex"
                /db_xref="taxon:9031"
                /clone="ChEST117j19"
                /dev_stage="16 day embryo"
                /lab_host="DRI08"
                /clone_lib="CSEQCHL24"
                /note="Organ: brain; Vector: pBluescript II KS(+); Site: 1:
                EcoRI; Site 2: NotI; Modification of pBluescript II KS(-)
                [Stratagene] vector to accommodate cDNA produced with the
                T-trimmed protocol (Construction of uni-directionally
                cloned cDNA libraries from messenger RNA for improved 3'
                end DNA sequencing by Glenn Fu, et al. U.S. Patent #
                6,387,624). Cut pBluescript II KS(+) with NotI and EcoRI.
                Ligate in double stranded adaptor containing BspI and
                BamHI sites [5'ggccgctg-cgagcccgatccgaaaaag]
                [5'aattcttttttcggatccgggtgcagc]"
ORIGIN
Query Match      56.0%; Score 692.8; DB 5; Length 860;
Best Local Similarity 89.6%; Pred. No. 3.9e-152;
Matches 756; Conservative 0; Mismatches 87; Indels 1; Gaps 1;
QY      81  CAGCAGCTAGCGTTGCCCACTTAGATCAACAATGGTAGACCAGCGATCAACTGTTT 140
DB      10  CAGCAGCTAGCGTTGCCCACTTAGATCAACAATGGTAGACCAGCGATCAACTGTTT 69
QY      141 TTCCTAAATGAAAGTGAATCGAAGACAGACTAAGACAACTGGAAACGCCCAAAAT 200
DB      70  TTCCTAAATGAAAGTGAATCGAAGACAGACTAAGACAACTGGAAACGCCCAAAAT 129
QY      201 GAACCTAGTGTGCTGGAAGTTTACCCCTGATACCAACACAGGGAAGTAAATGCGCAAG 260
DB      130 GAACCTAGTGTGCTGGAAGTTTACCCCTGATACCAACACAGGGAAGTAAATGCGCAAG 189
QY      261 TGTCGAATGCTTATCCAGGAGAATCAAGAGCTTGAAGGCGAGCTGTCCAGGGAGCTATT 320
DB      190 TGTCGAATGCTTATCCAGGAGAATCAAGAGCTTGAAGGCGAGCTGTCCAGGGAGCTATT 249

```

```

QY      321 GCACAACCTGAAGCAGAGTTGGCTTTTACAGAGAAATACAGTGGAGGCTTAAAGCAGT 380
DB      250 GCACAGCTTGAGGCAGAGTTGGCTTTTACAGAGAAATATAGTGGAGAACTTAAAGCAGT 309
QY      381 CAGATGAACCTGAATGACTTTCATCATCCAGCTTGATGAAGAAAGTAGAGGTTATCCAGAGT 440
DB      310 CAGATGAATGAATGACTTTCATCATCCAGCTTGATGAAGAAAGTAGAGGTTATCCAGAGT 369
QY      441 ACCATTCTAGTTCCTGACGACGACGCTGAAGAGACACGCCAGCAAGTTGGCTCAGTACCAG 500
DB      370 ACCATTCTAGTTCCTGACGACGACGCTGAAGAGAGACTCGCCAGCAGTTGGCGCAGTACCAG 429
QY      501 CAGCAGCAGTCTCAGGCTCTGCCCAAGTACCAGCAGGACTACAGCTTCTGAACTCTGTA 560
DB      430 CAGCAGCAGTCTCAGGCTCTGCCCAAGTACCAGCAGGACTACAGCTTCTGAGCCTTACA 489
QY      561 GAAACAGTCAGAGGCGCACAAAGTAAAGACTGCTCTGACAAACGACCAAGTAAATGGT 620
DB      490 GACCAG-CAGAGGCTGTGGGTAAGACTGACGCGCTCTGGCTAAACGACCAAGCAATGGT 548
QY      621 AGCTCTCTCCGCGCAGAGCAGCTCTGGGCTCTGGATTTACAGGAGGCGGCAACACACGAA 680
DB      549 AGCTCTCTCCGCGCAGAGCAGCTCTGGGCTCTGGATTTATAGGAGGCTGAGCAGCAGAA 608
QY      681 GATGACTTTCCTTCTTCTCCAGGGAATGGTAAATAGTCTCTCCAAACAGCTCAGAGGAGAGA 740
DB      609 GATGACTTTCCTTCTTCTCCAGGGAATGGTAAATAGTCTCTCCAAACAGCTCAGAGGAGAGA 668
QY      741 ACTGCGCAGAGGAGTGTGGTTAGCTAAATCAACTCAGTCTGGGGTATGAAAGTGTAGAC 800
DB      669 ACTGCTAGAGGAGTGTGGTTAGCTAAATCAACTCAGTCTGGGGTATGAAAGTGTAGAC 728
QY      801 TCTCCACGCGGCGAGTGAATACTCTCTCACACACCAATCAATGACACAGACTCCAGTCAAT 860
DB      729 TCTCCACGCGGCGAGTGAATACTCTCTCACACCAATCAATGACACAGACTCCAGTCAAT 788
QY      861 GACCTCTAAGAGGAGAGAAAGCAGTGTGGGAAAGTAATCGAATCTGTGGGTTCCGCGCAC 920
DB      789 GATCTCTAAGAGGAGAGAAAGCAGTGTGGGAAAGTAATCGAATCTGTGGGTTCCGCGCAT 848
QY      921 GTTC 924
DB      849 GTCC 852

```

```

RESULT 6
BX395380      1130 bp  mRNA  linear  EST 01-MAY-2004
LOCUS        BX395380 Homo sapiens NEUROBLASTOMA COT 50-NORMALIZED Homo sapiens
DEFINITION   cDNA clone CSDD0006YF24 5-PRIME, mRNA sequence.
ACCESSION    BX395380
VERSION      BX395380.2 GI:46920201
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 1130)
AUTHORS      Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE        Full-length cDNA libraries and normalization
JOURNAL      Unpublished (2001)
COMMENT      On May 13, 2003 this sequence version replaced gi:30628516.
              Contact: Genoscope
              Genoscope - Centre National de Sequencage
              BP 191 91006 EVRY cedex - France
              Email: seq-refgenoscope.cns.fr, Web : www.genoscope.cns.fr
              1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
              end enriched, double-strand cDNA was digested with Not I and cloned
              into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
              was normalized. Library was constructed by Life Technologies, a
              division of Invitrogen. This sequence belongs to sequence cluster
              3991.f

```

For more information about this cluster, see
<http://www.genoscope.cns.fr/cdna?c=CS0DD006DHL2QPl&c=3991.f>.

FEATURES

```

source
1..1130
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DD006VP24"
/tissue_type="NEUROBLASTOMA COT 50-NORMALIZED"
/clone_lib="Homo sapiens NEUROBLASTOMA COT 50-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-cligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Query Match      50.3%; Score 622.8; DB 5; Length 1130;
Best Local Similarity 90.8%; Pred. No. 1.2e-135;
Matches 652; Conservative 26; Mismatches 38; Indels 2; Gaps 2;

QY 1 ACATCCTTGGTAATCGACTAGCAACCAAGGACAGATATTCAAGAGTGTAAGTCAAA 60
DB 414 ACATCCTTGTAAATCGGACTAGCAACCAAGGACAGATGCAAGAGTGTAAGTCAAA 473
QY 61 TCCAGTACCTCAAGCAAGTCCAGAGCGTACGCTGTCGCCAACTTAGATCAACAATGGTAG 120
DB 474 TCCAGTACCTCAAGCAAGTTCAGAGCGCGGTGTCGCCAACTGAGATCAACAATGGTAG 533
QY 121 ACCAGAGCGATCAACTGTTTCTTAAATGAAAGGTGAAGTGGACAGACTAAAGACA 180
DB 534 ACCAGAGCGATCAACTGTTTCTTAAATGAAAGGTGAAGTGGACAGACTAAAGACA 593
QY 181 AACTGGAAACAGCCCAAAATGAATGAGTGCCTGGAAAGTTTACGCTGTAGCCAAACAG 240
DB 594 AACTGGAAACAGCCCAAAATGAATGAGTGCCTGGAAAGTTTACGCTGTAGCCAAACAG 653
QY 241 GGAATAAGTTAATGGCGAAGTGTGGAATGCTTATCCAGGAGAGATCAAGAGCTTGGAGGC 300
DB 654 GGAATAAGTTAATGGCGAAGTGTGGAATGCTTATCCAGGAGATCAAGAGCTTGGAGGC 713
QY 301 AGCTGTCCAGGGACGATTTCACAACCTTGAAGCAGAGTTGGCTTTACAGAAGAAATACA 360
DB 714 AGCCGTRCCAGGGACGATTTCACAACCTTGAAGCAGAGTTGGCTTTACAGAAGAAATACA 773
QY 361 GTGAGGAGCTTAAAGCAGTCAAGATGAATGATCACTTCAATCCAGCTTGGATGAAG 420
DB 774 GTWAGGAGCTTAAAGCAGTCAAGATGAATGATCACTTCAATCCAGCTTGGATGAAG 833
QY 421 AAGTAGAGGGTATGCGAGAGTACCACTTCTAGTTCTGCGAGCAGCAGCTG-AAGGAGACAGC 479
DB 834 AAGTAGAGGGTATGCGAGAGTACCACTTCTAGTTCTGCGAGCAGCAGCTG-AAGGAGAAAGC 893
QY 480 CAGCAGTTGGCTGATGACAGCAGCAGCTCTCAGGCTCTGCCCAAGTACCAGCAGG 539
DB 894 CAGCAGTTGGCTGATGACAGCAGCAGCTCTCAGGCTCTGCCCAAGTACCAGCAGG 952
QY 540 ACTACAGCTTCTGAACCTGTAGAACAGTCAAGGCGCACAGTAAAGACTGCACTGCTG 599
DB 953 ACTACAGCTTCTGAACCTGTAGAACAGTCAAGGCGCACAGTAAAGACTGCACTGCTG 1012
QY 600 ACAAACGGACAAGTAATGTAGTCTCTCCCGCAGGAGCAGCTCTGGGTCTGGATTTTAC 659
DB 1013 ASAAACGGACAAGTAATGTAGTCTCTCCCGCAGGAGCAGCTCTGGGTCTGGATTTTAC 1072
QY 660 AGGAGGGGCAACACACCGAAGTACTTTCCTTCTTCCAGGGGAATGTAATAGT 717
DB 1073 AGGAGGGGCAACACACCGAAGTACTTTCCTTCTTCCAGGGGAATGTAATAGT 1130

```

RESULT 7
 CN166400/c
 LOCUS
 DEFINITION 997672 MARC 4P1G Sus scrofa cDNA 3', mRNA sequence. EST 02-APR-2004

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

CN166400
 CN166400.1 GI:46180830
 EST.
 Sus scrofa (pig)
 Sus scrofa

REFERENCE
 AUTHORS

1 (bases 1 to 710)
 Smith,T.P.L., Fraking,B.A., Ford,J.J., Vallet,J.L., Wise,T.A.,
 Nonneman,D.J., Wray,J.E. and Keese,J.W.

TITLE

Porcine EST collection using a normalized library constructed from
 embryos representing early developmental stages

JOURNAL

COMMENT

USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390

Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called with phred v0.020425.c and
 trimmed with the aid of the trim_alt option. Vector identified with

cross_match v0.990329.

Placer: TMM8066 row: L column: 12

Seq primer: TAGAAGGCACAGTCGAGG.

Location/Qualifiers

1..710

/organism="Sus scrofa"

/mol_type="mRNA"

/db_xref="taxon:9823"

/tissue_type="pooled"

/lab_host="DHI0B"

/clone_lib="MARC 4P1G"

/note=Vector: pCDNA3.1; Site 1: EcoRI; Site 2: NotI;

Library made with combined RNA from day-10, day-13,

day-15, day-25, and day-30 whole embryos."

ORIGIN

Query Match 49.1%; Score 608; DB 7; Length 710;
 Best Local Similarity 91.9%; Pred. No. 3.4e-132;
 Matches 654; Conservative 0; Mismatches 55; Indels 3; Gaps 1;

QY 262 GTCGAATGCTTATCCAGGAGATCAAGAGCTTGAAGGACAGCTGCCAGGAGCTATTG 321
 DB 710 GTCGAATGCTTATCCAGGAGATCAAGAGCTTGAAGGACAGCTGCCAGGAGCTATTG 651
 QY 322 CACAACCTTGAAGCAGAGTGGCTTTTACAGAAGAAATACAGTGAGGAGCTTAAAGCAGTC 381
 DB 650 CACAACCTTGAAGCAGAGTGGCTTTTACAGAAGAAATATAGTGAGGAGCTTAAAGCAGTC 591
 QY 382 AGGATGAAGTGAAGTTCATCATCCAGCTTGTATGAAGTAGAGGGTATCCAGAGTA 441
 DB 590 AGGATGAAGTGAAGTTCATCATCCAGCTTGTATGAAGTAGAGGGTATCCAGAGTA 531
 QY 442 CCATTCTAGTTCAGCAGCAGCTGAAGGAGACACCCAGCAGCTTGGCTCAGTACCAGC 501
 DB 530 CCATTCTAGTTCCTTCAACAGCAGCTGAAGGAGACCCAGCAGCTTGGCTCAGTACCAGC 471
 QY 502 AGCAGCAGTCTCAGGCTCTGCCCAAGTACCAGCAGCTACAGCTTCTGACCTGTAG 561
 DB 470 AGCAGCAGTCTCAGGCTCTGCCCAAGTACCAGCAGCTACAGCTTCTGACCTGTAG 414
 QY 562 AACAGTCAGAGGCGCACAAAGTAAAGACTGCACTGCTGTGACAAACGGACCAAGTAATGGTA 621
 DB 413 GACAGGAGAGGCGCACAAAGTAAAGACTGCACTGCTGTGACCAACGGACCAAGTAATGGTA 354
 QY 622 GCTCCTCCCGCAGAGGAGCTCTGGGTCTGGATTTTACAGGGAGGGCAACACAAACCGAAG 681
 DB 353 GCTCCTCCCGCAGAGGAGCTCTGGGTCTGGGTCTGGGTCTGGGTCTGGGTCTGGGTCTGGGT 294
 QY 682 ATGACTTCTCTCTTCTCCAGGGAATGGTAATTAAGTCTCTCAACAGCTCAGAGGAGAGAA 741
 DB 293 ATGACTTCTCTCTTCTCCAGGGAATGGTAATTAAGTCTCTCAACAGCTCAGAGGAGAGAA 234

```

QY 742 CTGGCAGAGAGTAGTGGTTAGCTAAATCAACTCAGTCGCGGGTATGAAGTGTAGACT 801
Db 233 CTGGCAGAGAGTAGTGGTTAGCTAAATCAACTCAGTCGCGGGTATGAAGTGTAGACT 174
QY 802 CTCCACGGGCACTGAAACTCTCTCACACACCAATCAAAATGACACAGACTCCAGTCATG 861
Db 173 CTCCACGGGCACTGAAACTCTCTCACACGGCACTCACATGACACAGACTCCAACTCG 114
QY 862 ACCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 921
Db 113 ACCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 54
QY 922 TCCAGATGCTTGGACTCAAGTGTAAATGTACAGGGTTCAGTTTGTAAATA 973
Db 53 TCCAGATGCTTGGACTCAAGTGTAAATGTACAGGGTTCAGTTTGTAAATA 2

RESULT 8
BI089654
LOCUS 602854348F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:4996066 5',
DEFINITION mRNA sequence.
ACCESSION BI089654
VERSION BI089654.1 GI:14507984
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 758)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHAM1020 row: 0 column: 11
High quality sequence start: 8
High quality sequence stop: 622.
FEATURES
source
1..758
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4996066"
/cell_line="MGC36"
/lab_hosts="DH10B"
/clone_lib="NIH MGC 10"
/notes="Organ: cervix; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: oligo dt.
Average insert size 1.5 kb. Library prepared by Life
Technologies."

ORIGIN
Query Match 48.8%; Score 604.4; DB 4; Length 758;
Best Local Similarity 92.68; Pred No. 2.5e-131;
Matches 646; Conservative 0; Mismatches 51; Indels 1; Gaps 1;

QY 476 ACGCCAGAGTGGCTCAGTACACAGCAGCAGTCTCAGCGCTCTGCCCAAGTACCAG 535
Db 13 ACGCCAGAGTGGCTCAGTACACAGCAGCAGTCTCAGCGCTCTGCCCAAGTACCAG 72
QY 536 CAGGACTACAGTCTCAGTACCTGTAGAACAGTACAGGCCACAGTAAAGACTGCAGTCG 595
Db 73 CAGGACTACAGTCTCAGTACCTGTAGAACAGTACAGGCCACAGTAAAGACTGCAGTCG 132
QY 596 TCTGACAAACGACCAAGTAAATGGTAGCTCTCCCGCCAGAGGACGTCCTGGGTCTGGATT 655

```

```

Db 133 TCTGACAAACGACCAAGTAAATGGTAGCTCTCCCGCCAGAGGACGTCCTGGGTCTGGATT 192
QY 656 TCACAGGAGGAGGCAACCAACACCGAAGATGATCTTCTTCTTCTCAGGGAATGATATAA 715
Db 193 TCACAGGAGGAGGCAACCAACACCGAAGATGATCTTCTTCTTCTCAGGGAATGATATAA 252
QY 716 GTCTCTCAACAGCTCAGAGGAGAGAACTGGCAGAGGAGGTAGTGGTTACGTAATCAACT 775
Db 253 GTCTCTCAACAGCTCAGAGGAGAGAACTGGCAGAGGAGGTAGTGGTTACGTAATCAACT 312
QY 776 CAGTCGCGGGTATCAAAAGTGTAGACTCTCCACCGGCGAGTGAATACTCTCTCACACCCA 835
Db 313 CAGTCGCGGGTATCAAAAGTGTAGACTCTCCACCGGCGAGTGAATACTCTCTCACACCCA 372
QY 836 ATCAAAATCACACAGACTCCAGTCATGACCCCTCAAGAGGAGAGAAACAGTGTAGTGAAGG 895
Db 373 ATCAAAATCACACAGACTCCAGTCATGACCCCTCAAGAGGAGAGAAACAGTGTAGTGAAGG 432
QY 896 TAAATCGAACTGTGGGTTCCTCCCGCCACGTTCAAGATGGCTTGGACTCAAGTGTAAATGTACA 955
Db 433 TAAATCGAACTGTGGGTTCCTCCCGCCACGTTCAAGATGGCTTGGACTCAAGTGTAAATGTACA 492
QY 956 GGGTTCAGTTTGTAAATA-TTTTTTCAGCAAAATTTTATACAGTGTCAATTAATTCGGA 1014
Db 493 GGGTTCAGTTTGTAAATA-TTTTTTCAGCAAAATTTTATACAGTGTCAATTAATTCGGA 552
QY 1015 CAGGATCTGTCCAGAAAATTAATGCATACATTTTGTCACAATTTGCGCTTTTGTGGTGT 1074
Db 553 CAGGATCTGTCCAGAAAATTAATGCATACATTTTGTCACAATTTGCGCTTTTGTGGTGT 612
QY 1075 ACGTTTTCGTTTTTTTGTGTTTTTTTGTGTTTTTTTGTGTTTTTTTGTGTTTTTTTGTG 1134
Db 613 ACGTTTTCGTTTTTTTGTGTTTTTTTGTGTTTTTTTGTGTTTTTTTGTGTTTTTTTGTG 672
QY 1135 AGGTGGTGTGTTTTTTTGTGTTTTTTTGTGTTTTTTTGTGTTTTTTTGTGTTTTTTTGTG 1172
Db 673 AGGTAAATAAATAAAGAGGGGCAAAATTTGGAGC 710

CO387170 808 bp mRNA linear EST 30-JUN-2004
AGENCOURT 26625081 NIH_MGC_253 Rattus norvegicus cDNA clone
IMAGE:7302938 5', mRNA sequence.
CO387170
CO387170.1 GI:49494761
EST.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 808)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Drs. Josef Lazar & Howard Jacob, Medical
College of Wisconsin
cDNA Library Preparation: Express Genomics
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHAM15337 row: b column: 24
High quality sequence stop: 654.
Location/Qualifiers

```

```

source
1. 808
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="IMAGE:7302938"
/tissue_type="Ovary - Brown Norway Line 3 Age 8 weeks,
free the entire procedure."
/lab_host="DH10B Tona"
/clone_lib="NIH_MGC_253"
/notes="Organ: ovary; Vector: pExpress-1; Site:1: EcoRV;
Site 2: NotI; RNA obtained from female animals at 8 wk
old. Tissues were snap-frozen and kept at -80C for two
days before RNA extraction and purification (Tri-reagent
method). cDNA was primed using oligo-dT primer:
5'-pGACTAGTCTAGATCGGAGCGGCCGCC(T)25-3' and cloned into
the EcoRV/NotI sites of pExpress-1. Size-selection 1.25kb
resulted in an average insert size of 1.5 kb. This primary
library is normalized (non-normalized primary library is
NIH MGC 252) and was constructed by Express Genomics
(Fredrick, MD). Note: this is a NIH_MGC library."

ORIGIN
Query Match 48.7%; Score 602.6; DB 7; Length 808;
Best Local Similarity 89.6%; Pred. No. 6.6e-131;
Matches 683; Conservative 0; Mismatches 70; Indels 9; Gaps 3;

QY 255 GCGAAGTGTGCAATGCTTATCCAGGAGAAATCAAGAGCTTGGAAAGGAGCTGTCCAGGGA 314
DB 14 GCGAAGTGTGCAATGCTTATCCAGGAGAAATCAAGAGCTTGGAAAGGAGCTGTCCAGGGA 314
QY 315 GGTATTGACAACTTGAAGCAGAGTTCGGCTTTACAGAGAAATACAGTGGAGGCTTAAA 374
DB 74 GGTATTGACAACTTGAAGCAGAGTTCGGCTTTACAGAGAAATACAGTGGAGGCTTAAA 133
QY 375 AGCAGTCAGGATGAATGATGCTTATCATTCAGCTTATGATGAAGTGTAGGGTATG 434
DB 134 AGCAGTCAGGATGAATGATGCTTATCATTCAGCTTATGATGAAGTGTAGGGTATG 193
QY 435 CAGAGTACCATTCCTGAGCAGCAGCTGAGGAGACAGCCAGGAGTGGGTCTAG 494
DB 194 CAGAGTACCATTCCTGAGCAGCAGCTGAGGAGACAGCCAGGAGTGGGTCTAG 253
QY 495 TACAGCAGCAGAGCTCTAGGCTCTGCCCAAGTACCAGAGGACTACAGCTTCTGAA 554
DB 254 TACAGCAGCAGAGCTCTAGGCTCTGCCCAAGTACCAGAGGACTACAGCTTCTGAA 313
QY 555 CCTGTAGAACAGTCAGAGGCCACAAAGTAAAGCTGCAGTCTGTGACAAACGGACCAAGT 614
DB 314 CCTGTAGAACAGTCAGAGGCCACAAAGTAAAGCTGCAGTCTGTGACAAACGGACCAAGT 373
QY 615 AATGTTAGCTCTCCCGCCAGAGAGCTCTGGGTCTGGATTTACAGGAGGAGGACACA 674
DB 374 AATGTTAGCTCTCCCGCCAGAGAGCTCTGGGTCTGGATTTACAGGAGGAGGAGCACT 433
QY 675 ACCGAAGATGACTTTCCTCTCTCAGGGAATGGAATAAGTCTCCCAACAGCTCAGAG 734
DB 434 CTTGAGATGACTTTCCTCTCTCAGGGAATGGAATAAGTCTCCCAACAGCTCAGAG 493
QY 735 GAGAGAACTGGCAGAGAGTGTGTTACGTTAAATCAACTCAGTCGGGGTATGAAAGT 794
DB 494 GAGAGAACTGGCA---GAGGTAGTAGTTACATAAACCCACTCAGTCGGGGTATGAAAGT 550
QY 795 GTAGACTCTCCAGGAGGAGTGAATCTCTCACACACCAATCAATGACACAGCTCC 854
DB 551 GTAGACTCTCCAGGAGGAGTGAATCTCTCACACACCAATCAATGACACAGCTCC 610
QY 855 AGTCATGACCTCAAGAGGAGAGAGTGTGTTAGGGAAGGTAATCGAAGTCTGGGGTCC 914
DB 611 AGTCATGACCTCAAGAGGAGAGAGTGTGTTAGGGAAGGTAATCGAAGTCTGGGGTCC 670
QY 915 CGCAGCTTCAGATGGCTTGGACTCAAGTGAATGATGACAGGTTGAGTTTGAATAT 974

```

```

Db 671 CCCCCATGTCAGAAATGGCTTTGGACTCAAGTGT-AATGTACAGGGGTCAATTTGTATATTT 729
QY 975 TTTTTCAGCAAAATTTTATACAGTGTCAATTTAATTTGGGAGA 1016
Db 730 TTACAGC-----AATTTTATACAGTGTCAATTAATGGGGAAGA 766

RESULT 10
BU942029
LOCUS
DEFINITION
710 bp mRNA linear EST 18-OCT-2002
AGENCY: 10556311 NIH MGC 127 Homo sapiens cDNA clone
IMAGE: 6714512 5', mRNA sequence.
BU942029
ACCESSION
BU942029
VERSION
BU942029.1 GI:24130848
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 710)
NIH-MGC http://mgi.nci.nih.gov/
AUTHORS
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: NCI
cDNA Library Preparation: Michael Brownstein Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LICM3026 row: i column: 08
High quality sequence stop: 555.

FEATURES
Location/Qualifiers
1..710
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6714512"
/tissue_type="mixed (pool of 40 RNAs)"
/lab_host="DH10B (Ti-Phage-resistant)"
/clone_lib="NIH_MGC_127"
/notes="Vector: pDNR-LIB; Site 1: SfiI (ggccattatggcc);
Site 2: SfiI (ggccgctcgcc); Double-stranded cDNA was
prepared from a pool of 40 cell line polyA+ RNAs (bladder
- 2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon -
4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%,
kidney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell -
5.2%, ovary - 4%, pharynx - 2.5%, prostate - 4.3%,
salivary gland - 1.3%, and skin - 2.3%). 5' and 3'
adaptors were used in cloning as follows:
5'-AAGCAGTGTATCAACGACAGTGGCCATTCAGCCGCGG-3' and
5'-ATTCTAGAGCCGAGGCGGCAGATG-dt(30)NN-3'. Full-length
enriched library was constructed using the Clontech
Creator SMART kit and size-selected to contain the 1-2 kb
size fraction (other fractions present in NIH_MGC_126 and
NIH_MGC_128). Library created in the laboratory of T.
Usdin, M.D., Ph.D. (NIMH, NIH). Note: this is a NIH_MGC
Library."

ORIGIN
Query Match 48.1%; Score 595.6; DB 5; Length 710;
Best Local Similarity 97.3%; Pred. No. 2.9e-129;
Matches 638; Conservative 0; Mismatches 14; Indels 4; Gaps 3;

QY 231 AGCCAAACAGGGAAGTAAATGGGAGAGTGTGCAATGCTTATCCAGGAGATCAAGAG 290
DB 45 AGCCAAACAGGGAAGTAAATGGGAGAGTGTGCAATGCTTATCCAGGAGATCAAGAG 104
QY 291 CTTGGAGGAGGAGTGTCCAGGAGAGCTATGCAACTTGAAGCAGAGTTGGCTTACAG 350
DB 105 CTTGGAGGAGGAGTGTCCAGGAGAGCTATGCAACTTGAAGCAGAGTTGGCTTACAG 164

```



```

351 AAGAAATACAGTGGAGGCTTAAAGCAGTCAGGATGAATGACTTTCATCTCCAG 410
Db 165 AAGAAATACAGTGGAGGCTTAAAGCAGTCAGGATGAATGACTTTCATCTCCAG 224
QY 411 CTTGATCAAGAGTATGAGGGTATGACAGTATACATTTCTAGTTCTGACAGCAGCAGCTGAAG 470
Db 225 CTTGATCAAGAGTATGAGGGTATGACAGTATACATTTCTAGTTCTGACAGCAGCAGCTGAAG 284
QY 471 GAGACAGCAGCAGTGGTCTAGTACCAGCAGCAGCTCTCAGCCCTCTGCCCCAAGT 530
Db 285 GAGACAGCAGCAGTGGTCTAGTACCAGCAGCAGCTCTCAGCCCTCTGCCCCAAGT 344
QY 531 ACCAGCAGGACTACAGCTTCTGAACCTGTAGAACAGTCAGAGGCCCAAGTAAAGACTGC 590
Db 345 ACCAGCAGGACTACAGCTTCTGAACCTGTAGAACAGTCAGAGGCCCAAGTAAAGACTGC 404
QY 591 AGTCGTCTGACAAACGAGCAAGTAATGTAGTCTCTCCCGCAGAGGACGCTGGGTCT 650
Db 405 AGTCGTCTGACAAACGAGCAAGTAATGTAGTCTCTCCCGCAGAGGACGCTGGGTCT 464
QY 651 GGATTTCCAGAGGGGCAACACACCGAAGATGACTTTCTCTTCTCCAGGGAATGTT 710
Db 465 GGATTTCCAGAGGGGCAACACACCGAAGATGACTTTCTCTTCTCCAGGGAATGTT 524
QY 711 AATAAGTCTCCACAGCTCAGAGGAGAGAACTGGCAGAGGAGTGTAGTGTACGTAAT 770
Db 525 AATAAGTCTCCACAGCTCAGAGGAGAGAACTGGCAGAGGAGTGTAGTGTACGTAAT 584
QY 771 CAACTGATGGGGTATGAAGTGTAGACTCTCCACGGGCAGT-GAATACTCTCTCAC 829
Db 585 CAACTGATGGGGTATGAAGTGTAGACTCTCCACGGGCAGTGTAGTGTACGTAAT 644
QY 830 AC-ACCAATCAATGACACAGA--CTCCAGTCTATGACCCCTCAAGAGGAGAGAAAG 882
Db 645 ACACAATCAATGACACAGA--CTCCAGTCTATGACCCCTCAAGAGGAGAGAAAG 700

RESULT 11
CB419819/5 656 bp mRNA linear EST 25-MAR-2003
LOCUS 592705 MARC 6BOV Bos taurus cDNA 3', mRNA sequence.
DEFINITION CB419819
ACCESSION CB419819.1 GI:29185262
VERSION CB419819.1
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovinae; Bos.
REFERENCE 1 (bases 1 to 656)
AUTHORS Smith,T.P.L., Roberts,A.J., Ehternkamp,S.E., Chitko-McKown,C.G.,
Wray,J.E. and Keefe,J.W.
TITLE A second set of bovine ESTs from pooled-tissue normalized libraries
JOURNAL Unpublished (2003)
COMMENT Contract: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and
trimmed with the aid of the trim_alt option. Vector identified with
cross_match v0.990329.
Plate: FQ8013 row: J column: 10
Seq primer: TAGAAGGCACAGTCGAGG.
Location/Qualifiers
1 .656
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"
/lab_host="DH10B"

FEATURES
source

```

```

/clone.lib="MARC 6BOV"
/notes=vector; pcDNA3.1; Site 1: EcoRI; Site 2: NotI;
Library made with RNA pooled from multiple tissues
including liver, lung, hypothalamus, pituitary, and
placenta/endometrium."

Query Match 48.0%; Score 594.2; DB 6; Length 656;
Best Local Similarity 94.2%; Pred. No. 6e-129;
Matches 517; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 217 AGTTTACCCCTGATAGCCAAACAGGGAAAAGTTAATGGGGAGTGTGGAATGTTATCC 276
Db 656 AGTTTACCCCTGATAGCCAAACAGGGAAAAGTTAATGGGGAGTGTGGAATGTTATCC 597
QY 277 AGGAAATCAAGAGCTTGGAAAGGAGCTGTCCAGGAGCGTATTGCGACAACTTGAAGCAG 336
Db 596 AGGAAATCAAGAGCTTGGAAAGGAGCTGTCCAGGAGCGTATTGCGACAACTTGAAGCAG 537
QY 337 AGTTGGCTTTACAGAAGAAATACAGTGGAGGCTTTAAAGCAGTCAGATGAATG 396
Db 536 AGTTGGCTTTACAGAAGAAATATAGTGGAGAGCTTTAAAGCAGTCAGATGAATG 477
QY 397 ACTTCATCATCCAGCTTGCATGAAGAAGTAGAGGGTATGACAGTACCATTTAGTCTTCG 456
Db 476 ACTTCATCATTCACCTTGACGAAGAAGTAGAGGGTATGACAGTACCATTTAGTCTTCG 417
QY 457 AGCAGCAGCTGAAGGAGACACGCCAGCAGTTGGCTCAGTACCGACGAGCAGTCTCAGG 516
Db 416 AGCAACAACTGAAGGAGACGCCAGCAGTTGGCTCAGTACCAACAGCAGCAGTCTCAAG 357
QY 517 CCTCTGCCCAAGTACAGCAGGAGTACAGCTTCTGACCTGTAGACCTGTAGACAGTCAGAGGCCA 576
Db 356 CGCAGGCCCCAGTACCAGCAGGACTACATCTCGAGCCCTGTAGGACGAGCAGAGGCCA 297
QY 577 CAAGTAAAGAGCTGCAGTCTCTGACAAACGAGCAAGTAATGGTAGTCTCTCCGCCCAGA 636
Db 296 CAGGTAAGAGCTGCAGTCTCTGSCCAACGAGCAAGTAATGGCAGTCTCTCCGCCCAGA 237
QY 637 GGACCTCTGGTCTGGATTTACAGGGAGGGCAACACACACCGAAGATGACTTCTCTTT 696
Db 236 GGACCTCTGGTCTGGATTTACAGGGAGGGGGACACAGCCGGAAGATGACTTCTCTTT 177
QY 697 CTCAGGGAATGGTAATAAGTCTCTCAACAGCTCAGAGGAGAGAACTGGCAGAGGAGGTA 756
Db 176 CTCAGGGAATGGTAATAAGTCTCTCAACAGCTCAGAGGAGAGAACTGGCAGAGGAGGTA 117
QY 757 GTGGTTACGTAATAACAACTCAGTGGGGTATGAAAGTGTAGACTCTCCACGGGCACTG 816
Db 116 GTAGTTAGCTAAACCAACTCAGTGGGGTATGAAAGTGTAGACTCTCCACGGGCACTG 57
QY 817 AAAAATCTCTCACACACCACTCAATGACAGACTCCAGTCCATGACCTCAAGA 871
Db 56 AAAAATCTCTCACACACCACTCAATGACAGACTCCAGTCCATGACCTCAAGA 2

RESULT 12
BI185059/5 702 bp mRNA linear EST 10-JUL-2001
LOCUS UNL-P-FN-cl-a-02-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
DEFINITION UNL-P-FN-cl-a-02-0-UNL 3', mRNA sequence.
ACCESSION BI185059
VERSION BI185059.1 GI:14659468
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 702)
AUTHORS Caetano,A.R., Johnson,R.K. and Pomp,D.
TITLE Generation and sequence characterization of a normalized cDNA
JOURNAL library from swine ovarian follicles
Mamm. Genome 14 (1), 65-70 (2003)

```

MEDLINE 22419904
 PUBMED 12532269
 COMMENT Contact: Pomp, D
 Department of Animal Science
 University of Nebraska, Lincoln
 Lincoln, NE 68583-0908, USA
 Tel: 402 472 6416
 Fax: 402 472 6362
 Email: dpomp@unl.edu
 Oligo-dT track not found, Not 1 site shown in beginning of sequence
 is likely internal to the message.
 Seq primer: M13-29
 POLYA=No.

FEATURES
 source
 1..702
 /organism="Sus scrofa"
 /mol_type="mRNA"
 /strain="University of Nebraska, Lincoln Swine Selection
 Lines"
 /db_xref="taxon:9823"
 /clone="UNL-P-FN-cl-a-02-0-UNL"
 /dev_stage="ADULR"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="UNL-P-FN"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not 1; Site 2: Eco RI; The UNL-P-FN
 library is a normalized library representing porcine
 ovarian follicles, ranging between 2.0 to 10.0 mm in
 diameter, collected during 7 days of the follicular phase
 of the pig estrous cycle. This library was derived from
 the library UNL-P-F2. The tag is a string of 5-6
 nucleotides present between the Not I site and the
 oligo-dT track. The library was constructed as described
 by Ronaldo, Lennon and Soares, Genome Research 6:
 791-806, 1996.
 TAG_Seq=None found"

ORIGIN
 Query Match 47.7%; Score 590.6; DB 4; Length 702;
 Best Local Similarity 91.6%; Pred. No. 4.3e-128;
 Matches 636; Conservative 0; Mismatches 55; Indels 3; Gaps 1;

QY 249 TTAATGGCGAAGTGTGCAATGCTTTATCCAGAGAAATCAAGAGCTTGGAGGCGAGTGTCC 308
 DB |||||
 QY 702 TTAATGGCGAAGTGTGCAATGCTTTATCCAGAGAAATCAAGAGCTTGGAGGCGAGTGTCC 643
 DB |||||
 QY 309 CAGGAGCTATTGCACAACTTGAAGCAGAGTGGCTTTACAGAGAAATACAGTCAGGAG 368
 DB |||||
 QY 642 CAGGAGCTATTGCACAACTTGAAGCAGAGTGGCTTTACAGAGAAATATAGTAGGAG 583
 DB |||||
 QY 369 CTTAAAGCAGTCAGGATGAATGACTTATCATCATCAGCTTGTATGAAGAAATAGAG 428
 DB |||||
 QY 582 CTTAAAGCAGTCAGGATGAATGACTTATCATCATCAGCTTGTATGAAGAAATAGAG 523
 DB |||||
 QY 429 GGTATCGAGATGACCTTCTAGTTCGACGACGCTGAAGGAGACGCCAGCAGTTG 488
 DB |||||
 QY 522 GGTATCGAGATGACCTTCTAGTTCGACGACGCTGAAGGAGACGCCAGCAGTTG 463
 DB |||||
 QY 489 GCTCAGTACGAGCAGCAGCTCTCAGGCTCTGCCCAAGTACCAGCAGGAGTACAGCT 548
 DB |||||
 QY 462 GCTCAGTACGAGCAGCAGCTCTCAGGCTCTGCCCAAGTACCAGCAGGAGTACAGCT 406
 DB |||||
 QY 549 TCTGACCTGTAGAACAGTCAGAGGCCACAAAGTAAGACTGAGTCGTCTGACAAACGGA 608
 DB |||||
 QY 405 TCTGACCTGTAGAACAGTCAGAGGCCACAAAGTAAGACTGAGTCGTCTGACAAACGGA 346
 DB |||||
 QY 609 CCAAGTAAGTGTAGCTCTCTCCGCCAGAGAGCTCTGGGTCTGGATTTTACAGGAGGGC 668
 DB |||||
 QY 345 CCAAGTAAGTGTAGCTCTCTCCGCCAGAGAGCTCTGGGTCTGGATTTTACAGGAGGGC 286
 DB |||||
 QY 669 AACACAAACGAGATGACTTCTCTTCTCAGGGAATGGTAAGTCTCTCCAAACAGC 728
 DB |||||
 QY 285 GACACACCGAAGATGACTTCTCTTCTCCGGGAGCGGAATAGGCTCTCCAAACAGC 226
 DB |||||

QY 729 TCAGAGGAGAGAACTGGCAGAGGAGGTAGTGTACGTAAATCAACTCAGTGGCGGTAT 788
 DB |||||
 QY 225 TCAGAGGAGAGAACTGGCAGAGGAGGTAGTGTACGTAAATCAACTCAGTGGCGGTAT 166
 DB |||||
 QY 789 GAAAGTGTAGACTCTCCACCGGCGAGTGAAGAACTCTCTCACACCAATCAATGACACA 848
 DB |||||
 QY 165 GAAAGCTGTAGACTCTCCACCGGCGAGTGAAGAACTCTCTCACACCAATCAATGACACA 106
 DB |||||
 QY 849 GACTCCAGTGTAGACTCTCAAGAGGAGAAAGCAGTGTAGTGGAAAGGTAAATCGAACTGTG 908
 DB |||||
 QY 105 GACTCCAACTCTGACCTCAAGAGGAGAAACCGTGTAGTGGAAAGGTAGCCGAACTGCG 46
 QY 909 GGTTCCTCCGCGAGTTCAGAAATGGCTTGGACTCAA 942
 DB |||||
 QY 45 GGTTCCTCCGCGAGTTCAGAAATGGCTTGGACTNCA 12
 DB |||||

RESULT 13
 BU200958
 LOCUS
 DEFINITION 880 bp mRNA linear EST 25-NOV-2002
 603949240F1 CSEQCHN03 Gallus gallus cdna clone CHEST906p9 5', mRNA
 sequence.
 ACCESSION BU200958
 VERSION BU200958.1 GI:25364439
 KEYWORDS EST
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 REFERENCE 1 (bases 1 to 880)
 Boardman,P.E., Sanz-Esquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
 Pong,W.F., Rickie,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
 A Comprehensive Collection of Chicken cDNAs
 Curr. Biol. 12 (22), 1965-1969 (2002)
 22335534
 MEDLINE
 PUBMED 12445392
 COMMENT Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology
 (UMIST)
 PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.

FEATURES
 source
 1..880
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="White Leghorn, Hisex"
 /db_xref="taxon:9031"
 /clone="CHEST906p9"
 /tissue_type="whole embryo"
 /dev_stage="20-21"
 /lab_host="DH10B"
 /clone_lib="CSEQCHN03"
 /note="Organ: whole embryo; Vector: pBluescript II KS(+);
 Site 1: EcoRI; Site 2: NotI; This normalized library was
 constructed from 1 million independent clones cDNA
 synthesis was initiated using an oligo(dT) primer, using
 methylated C in the first strand synthesis reaction.
 Following this first strand reaction, double-stranded cDNA
 was bluntended, ligated to NotI adapters, digested with
 EcoRI, size-selected, and cloned into the NotI and EcoRI
 compatible sites of a custom modified MCS of the
 pBluescript (KS+) vector. The library was normalized in 2
 rounds using conditions adapted from Soares et al., PNAS
 (1994) 91: 9228-9232 and Bernaldo et al., Genome Research 6
 (1996): 791, except that a significantly longer
 reannealing hybridization was used."

ORIGIN

Query Match	46.2%	Score 571.8;	DB 5;	Length 880;
Best Local Similarity	84.2%	Pred. No. 1.2e-123;		
Matches 656;	Conservative 0;	Mismatches 122;	Indels 1;	Gaps 1;
Qy	273	ATCCAGAGAGAAATCAAGAGCTTGAAGGACAGCTCTCCAGGAGCGTATTGCACAACTTGAA	332	
Db	1	ATCCAGAGAGAAATCAAGAGCTTGAAGGACAGCTCTCCAGGAGCGTATTGCACAGCTTGAG	60	
Qy	333	GCAGAGTGGCTTTACAGAGAAATACAGTGAGGAGCTTAAAGCAGTACAGGATGAAC	392	
Db	61	GCAGAGTGGCTTTACAGAGAAATACAGTGAGGAGCTTAAAGCAGTACAGGATGAATG	120	
Qy	393	AATGACTTTCATCCAGCTTGATGAAGAAGTAGAGGTTATGCAGAGTACCACTTCAGTT	452	
Db	121	AATGACTTTCATCCAGCTTGATGAAGAAGTAGAGGTTATGCAGAGTACCACTTCAGTT	180	
Qy	453	CTGCAGAGAGCTGAAGGAGACACGCCAGCAGTTGGCTCAGTACCAGCAGCAGCTCT	512	
Db	181	CTTCAGCAGCAGTTGAAGGAGACTCGCCAGCAGTTGGCGCAGTACCAGCAGCAGCTCC	240	
Qy	513	CAGGCTCTGCCCAAGTACCAGCAGGACTACAGCTTCTGAACCTGTAGAACAGTACAGAG	572	
Db	241	CAGGCTCTCCACCCAGGTACCAGCAGGACTCCATCTTCTGAGCTACAGACCCAGGGAG	300	
Qy	573	GCCAAAGTAAGACTCGCTGTCTGACAAACGACCAAGTAATGCTAGCTCTCTCCCGC	632	
Db	301	GCTGTGGTTAAGACTCAGCGCTGTGGCTAAACGCAAGCAATGGTAGCTCTCTCCCAT	360	
Qy	633	CAGAGGACGCTGGGTCTGGATTTCAGAGGAGGCAACACACCGAAGATGACTTTCCT	692	
Db	361	CAGCGGACGCTGGGCTGGATTTCAGAGGAGGTAAGCAGACAGAGATGACTTTCCTCG	420	
Qy	693	TCCTTCTCCAGGAAATGTAATAGTCTCTCAACAGCTCAGAGGAGAGAACTGCGCAGGA	752	
Db	421	GCTTCTCCAGGAAATGTAATAGTCTCTCAACAGCTCAGAGGAGAGAACTGCGTAGAG	480	
Qy	753	GGTAGTGGTTACGTAATCAACTCAGTTCGGGGTATCAAAAGTGTAGACTCTCCACGGC	812	
Db	481	GGTAGTGGTTACGTAATCAACTCAGTTCGGGGTATCAAAAGTGTAGACTCTCCACGGC	540	
Qy	813	AGTGAAACTCTCTCACACCAATCAATGACACAGACTCCAGTCAATGACCTTCAAGAG	872	
Db	541	AGTGAAACTCTCTCACACCAATCAATGACACAGACTCCAGTCAATGACCTTCAAGAG	600	
Qy	873	GAGAAAGCAGTGTGGGAAAGGTAATCGAACTGTGGGTTCCGCCACGTTTCAAGATGC	932	
Db	601	GAGAAACAGTAGATGAAAGGTAATCGAACTGTGGGTTCCGCCATGTCCAGAAATGGT	660	
Qy	933	TTGACTCAAGTGAATGTACAGGTTTCAGTTTGTAAATTTTTTCAGCAATTTTA	992	
Db	661	TTGACTCAAGTGAATGTACAGGTTTCAGTTTGTAAATTTTTTCAGCAATTTTAT	720	
Qy	993	TACAGTGTCAATTAATTTGGAGA-GGATAGTCTCCAGAAATTAATGCAATCTTTGT	1050	
Db	721	ACAGTGTCAATTAATAGGAGAGGAGTACTGTTCAGAAAGCGGTTATTACATAGT	779	
RESULT 14				
AJ657839		660 bp	mRNA	linear
LOCUS	AJ657839	KN277	Sus scrofa	CDNA clone C0005211_H09, mRNA sequence.
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				

COMMENT	Contact: Anderson SI Genomics and Bioinformatics Roslin Institute Roslin, Midlothian, EH25 9PS, UNITED KINGDOM Single pass sequencing. Bases called and trimmed with phred v0.020425.c. Vector identified by cross match with the -minscore 20 and -minmatch 12 options. Vector:pBlueScriptII (SK+) R. Site1: EcoRI R. Site2: NotI 5' Seq Primer M13F Normalised library constructed from pooled early embryos, from 8- cell stage to blastocysts. Clones available from UK Centre for Functional Genomics in Farm Animals, Roslin Institute, Roslin, Midlothian, UK, EH25 9PS, www.arkgenomics.org. Location/Qualifiers 1..660 /organism="Sus scrofa" /mol_type="mRNA" /db_xref="taxon:9823" /clone="C0005211_H09" /tissue_type="embryo" /clone_lib="KN277" /note="Vector: pBlueScriptII (SK+); Site 1: EcoRI; Site 2: NotI; Single pass sequencing. Normalised library constructed from pooled early embryos, from 8-cell stage to blastocysts."			
FEATURES				
source				
ORIGIN				
Query Match	45.8%;	Score 566.4;	DB 1;	Length 660;
Best Local Similarity	92.5%;	Pred. No. 2.1e-122;		
Matches 607;	Conservative 0;	Mismatches 46;	Indels 3;	Gaps 1;
QY	200	TGAACGTAGTGGCTGGAAGTTTACGCTGTATGCCAAACAGGAAAAAGTTAATGGCGAA	259	
Db	4	TGAACGTAGTGGCTGGAAGTTTACGCTGTATGCCAAACAGGAAAAAGTTAATGGCGAA	63	
QY	260	GTGTGCAATGCTTATCCAGGAGAAATCAAGAGCTTGAAGGACGCTGTCCAGGAGCAGCTAT	319	
Db	64	GTGTGCAATGCTTATCCAGGAGAAATCAAGAGCTTGAAGGACGCTGTCCAGGAGCAGCTAT	123	
QY	320	TGCACAACCTTGAAGCAGAGATTGGCTTTACAGAGAAATACAGTAGGAGGCTTAAAGCAG	379	
Db	124	TGCACAACCTTGAAGCAGAGATTGGCTTTACAGAGAAATATAGTAGGAGGCTTAAAGCAG	183	
QY	380	TCAGGATGAACCTGAATGACTTTCATCATCCAGCTTGATGAAGAAAGTAGAGGGTATCGAGAG	439	
Db	184	TCAGGATGAACCTGAATGACTTTCATCATTCACAGTTCAGAGAAAGTAGAGGGTATCGAGAG	243	
QY	440	TACCAATCTAGTTCTGCAGCAGCAGCTGGAAGAGACACGCCAGCAGTTGGCTCAGTACCA	499	
Db	244	TACCAATCTAGTTCTGCAGCAGCAGCTGGAAGAGACACGCCAGCAGCTGGCTCAGTACCA	303	
QY	500	GCAGCAGCAGTCTCAGGGCTCTGCCCAAGTAGTACCAGCAGGACTACAGCTTCTGAACCTGT	559	
Db	304	ACAGCAGCAGTCTCAGGGCTCTGCCCAAGTAGTACCAGCAGGACTACAGCTTCTGAACCTGT	360	
QY	560	AGAACAGTCCAGAGGCGCACAAAGTAAAGACTGCAAGTCTGTGACAAACGGACCAAGTAATGG	619	
Db	361	AGAACAGTCCAGAGGCGCACAAAGTAAAGACTGCAAGTCTGTGACCAACGGACCAAGTAATGG	420	
QY	620	TAGTCTCTCCGCCAGAGAGCGTCTGGGTCTGGATTTCACAGGAGGCGCAACACACCGA	679	
Db	421	CAGTCTCTCCGCCAGAGAGCGTCTGGGTCTGGATTTCACAGGAGGCGCGACACACCGA	480	
QY	680	AGATGACTTTCCTTCTTCTCCAGGGAATGTAATAGTCTCCAAACAGCTCAGAGGAGAG	739	
Db	481	AGATGACTTTCCTTCTTCTCCGGGAGCGGGAATAGGCCCTCCAAACAGCTCAGAGGAGAG	540	
QY	740	AATGCGCAGAGAGGTAGTGGTTACGTTAAATCAACTCAGTTCGGGGGTATGAAAGTGTAGA	799	
Db	541	AATGCGCAGAGAGGTAGTGGTTACGTTAAATCAACTCAGTTCGGGGGTATGAAAGTGTAGA	600	
QY	800	CTCTCCAGCGGCGAGTGAAACTCTCTCACACCAATCAAAATGACACAGACTCCA	855	
Db	601	CTCTCCAGCGGCGAGTGAAACTCTCTCACACCAATCAAAATGACACAGACTCCA	656	

```

RESULT 15
CB182874
LOCUS      CB182874               961 bp      mRNA      linear      EST 16-MAY-2003
DEFINITION AGNCOURT 11381782 NIH_MGC_164 Mus musculus cDNA clone
IMAGE:30213672 5', mRNA sequence.
ACCESSION  CB182874
VERSION     CB182874.2  GI:29133999
KEYWORDS   EST.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  NIH-MGC http://mgi.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            On Jan 30, 2003 this sequence version replaced gi:28181082.
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: Dr. David Rowe and Dr. Mina
            cDNA Library Preparation: Invitrogen Corp
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDAW0315 row: f column: 01
            High quality sequence start: 50
            High quality sequence stop: 696.
            Location/Qualifiers
                1..961
                    /organism="Mus musculus"
                    /mol_type="mRNA"
                    /db_xref="taxon:10090"
                    /clone="IMAGE:30243672"
                    /lab_host="DH10B (phage-resistant)"
                    /clone_lib="NIH_MGC_164"
                    /note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI;
                    Non-normalized full-length enriched library from pooled
                    mouse embryonic limb, maxilla and mandible, day 10.5 and
                    11.5 (size selected for the 0.5-1 kb fragments). Cloned
                    directionally, priming method: Oligo-dr. cDNA enrichment:
                    >1k bp, Average insert size 1.8k bp. Priming sequence:
                    5'-GACTAGTCTAGATCGGCGGCCCTTT-3'. Tissue contributed
                    by, David Rowe. Library constructed by ResGen, Invitrogen
                    Corp."
FEATURES             source
    source
ORIGIN
Query Match      44.7%; Score 553.4; DB 6; Length 961;
Best Local Similarity 85.5%; Pred. No. 2.6e-119;
Matches 678; Conservative 0; Mismatches 76; Indels 39; Gaps 4;

QY  447  CTAGTCTTCGACGAGCAGCTGACGAGACACGCGCAGCAGTTGGCTCAGTACCAGCAGCAG 506
Db    |||
QY  3    CTGGTACCGGTCCGGAATCTTCGGGATGAGCGACAGCGTTGGCATTCCCGACGACGAG 62
Db    |||
QY  507  CAGTCTCAGGCCTCTGCCCAAGTACCAGCAGGAGCTACAGCTTCTGAACTGTAGAACAG 566
Db    |||
QY  63    CAGTCTCAAGCTTCAGCTCCCAAGTACCAGCAGGAGCTACATCTTCTCAACCTGTAGATCAG 122
Db    |||
QY  567  TCAGAGGCCACAGTAAGAGCTCAGTCTGTACAAACGGACCAAGTAATGGTAGCTCC 626
Db    |||
QY  123  GCAGAGGTCAAGACGAAGAGCTCAGTCTGTGGCAATGGACCAAGTAATGGCAGCTCC 182
Db    |||
QY  627  TCCCGCCAGAGGAGCGTCTGGGCTGTGATTTTCACAGGAGGCGCAACACACCGAAGATGAC 686
Db    |||
QY  183  TCCCGCCAGAGGAGCGTCTGGGCTGTGATTTTCACAGGAGGCGGAGCACCTCAGGATGAC 242
Db    |||
QY  687  TTTCCTTCTCTCCAGGGAATGGTAATAGTCTTCCAAAGCTCAGGAGAGAACTGGC 746
Db    |||
QY  243  TTTCCTTCTCTTCAGGGAATGGTAATAGTCTTCCAAAGCTCAGGAGAGAACTGGC 302

```

```

QY  747  AGAGGAGGTAGTGGTTAAGTAAATCAACTCAGTCGGGGGTATGAAAGTGTAGACTCTCCC 806
Db    |||
QY  303  AGAGGATGTAGTAGTTACATAAACCCACTCAGTCGGGGGTATGAAAGTGTAGACTCTCCC 362
Db    |||
QY  807  ACGGGCAGTGAAGAACTCTCTACACACCAATCAATGACACAGACTCCAGTCATGACCCCT 866
Db    |||
QY  363  ACGGGCAGTGAAGAACTCTCTACACATCACTCAATGACACAGACTCCAGTCATGACCCCT 422
Db    |||
QY  867  CAAGAGGAGAAAGCAGTCAAGTGGGAAAGTAATCGAACTGTGGGTTCGCCGCACGCTTCAG 926
Db    |||
QY  423  CAAGAGGAGAAAGCAGTCAAGTGGGAAAGTAATCGAACTGTGGGTTCGCCGCATGTTTCAG 482
Db    |||
QY  927  AATGGCTTGGACTCAAGTGTAAATGATACAGGCTTCAGTTTGTGTAATATTTTTCAGCAAA 986
Db    |||
QY  483  AATGGCTTGGACTCAAGTGTAAATGATACAGGCTTCAGTTTGTGTAATATTTTTCAGAAAA 542
Db    |||
QY  987  TTTTATACAGTGTCAATTTAATTTGGGAGAGATCTGCCAGAAAAATTAATGCATACTT 1046
Db    |||
QY  543  TTTTATACAGTGTCAATTTAATTTGGGAAAGATCTGT-CAAAAAATTAATGCATACTT 601
Db    |||
QY  1047  TTGTCACAATTTGCCCTTTTGTGGGTGACGTTTGGTTTTTTTGTGTTTTTTTCT 1106
Db    |||
QY  602  TTGTCACAATTTACCTTTTGTGGGTGT--GTTTTTGTGTTTGTGTTTCTTTTTTTT 659
Db    |||
QY  1107  TTGGATTTTTTTTCTTTCTTTTAAAGAGGTGGTGTGTTGTTTTTGTCTCAATACTTC 1166
Db    |||
QY  660  TT-----TTTTTAATGCTTCAATACCTC 694
Db    |||
QY  1167  TCGCGGCTTGGGAATGTGAACAGTTAATTTGAATGTTGC-AAAAGGCAATTTTGTG 1225
Db    |||
QY  685  TGCCACTTTGAAAAATTGTAAACAGTTAATTTACTTTGAACATTGCTAAAAGGCAATTTTGTG 744
Db    |||
QY  1226  GAGGTCACAGTTA 1238
Db    |||
QY  745  TAGGGTCAAGGTA 757
Db    |||

```

Search completed: November 14, 2004, 14:19:01
Job time : 4409 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2004, 12:15:48 ; Search time 191 Seconds
(without alignments)
960.966 Million cell updates/sec

Title: US-10-030-389-2
Perfect score: 1608
Sequence: 1 MRLATKGQDIQECTTQIQL.....GSRHVNGLDSSVNVQGSVL 319

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Uniprot 02.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1593	99.1	396	1	WTAP_HUMAN	Q15007
2	1593	99.1	396	2	AAH69192	AAH69192
3	1531	95.2	396	1	WTAP_MOUSE	Q98696
4	1296.5	80.6	393	2	Q6P4K5	Q6P4K5
5	1296.5	80.6	393	2	AAH63362	AAH63362
6	1146.5	71.3	423	2	Q7SXL7	Q7SXL7
7	528.5	32.9	412	2	Q7KA73	Q7KA73
8	528.5	32.9	539	2	Q9Y091	Q9Y091
9	525.5	32.7	412	2	Q8ML74	Q8ML74
10	525.5	32.7	536	2	Q9V6T8	Q9V6T8
11	518	32.2	272	2	Q7Q0B0	Q7Q0B0
12	357	22.2	74	2	Q9BZS4	Q9BZS4
13	282	17.5	330	2	Q6PNZ9	Q6PNZ9
14	218	13.6	267	2	Q6PNZ9	Q6PNZ9
15	218	13.6	267	2	Q7R2P6	Q7R2P6
16	190.5	11.8	554	2	AAH07458	AAH07458
17	172.5	10.7	617	2	Q815B2	Q815B2
18	168	10.4	652	2	AAH38587	AAH38587
19	168	10.4	652	2	Q9K2U9	Q9K2U9
20	166	10.3	524	2	Q9KJ33	Q9KJ33
21	165.5	10.3	576	2	Q9KJ33	Q9KJ33
22	163	10.1	114	2	Q7Q0B1	Q7Q0B1
23	160	10.0	943	2	Q6TPG6	Q6TPG6
24	158.5	9.9	692	2	AAQ94949	AAQ94949
25	158.5	9.9	692	2	Q6TPG6	Q6TPG6
26	158.5	9.9	725	2	Q8PFY1	Q8PFY1
27	158.5	9.9	725	2	CAH44152	CAH44152
28	158.5	9.9	1509	2	Q7JN47	Q7JN47
29	158.5	9.9	1509	2	CAF31488	CAF31488
30	158.5	9.9	1888	2	Q7JM45	Q7JM45
31	158.5	9.9	1888	2	CAF31490	CAF31490

32	158.5	9.9	1927	2	Q7JMA6	Q7JMA6
33	158.5	9.9	1927	2	CAF31489	CAF31489
34	158.5	9.9	2048	2	Q7JMA2	Q7JMA2
35	158.5	9.9	2048	2	CAF31493	CAF31493
36	158.5	9.9	2101	2	Q7JMA3	Q7JMA3
37	158.5	9.9	2101	2	CAF31492	CAF31492
38	158.5	9.9	2134	2	Q7JMA4	Q7JMA4
39	158.5	9.9	2134	2	CAF31491	CAF31491
40	157	9.8	995	2	Q9NJ19	Q9NJ19
41	157	9.8	1229	2	Q9NJ22	Q9NJ22
42	157	9.8	1243	2	Q9NJ21	Q9NJ21
43	157	9.8	1253	2	Q9NJ20	Q9NJ20
44	157	9.8	1951	2	Q17042	Q17042
45	156.5	9.7	516	1	P54_ENTFC	P13692

ALIGNMENTS

RESULT 1
WTAP_HUMAN STANDARD; PRT; 396 AA.
ID WTAP_HUMAN Q98YJ7; Q9H4E2;
AC Q15007; Q98YJ7; Q9H4E2;
DT 15-JUL-1998 (Rel. 36, Created)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE WTAP_HUMAN (Human)
DE pre-mRNA splicing regulator female-lethal (2D) [FL(2D)]
DE Name=WTAP; Synonyms=KIAA0105;
GN Homo sapiens (Human)
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Ortega A., Penalva L., Valcarcel J.,
RT "CDNA of human homolog of the drosophila putative pre-mRNA splicing
RT regulator female-lethal (2D) [FL(2D)]";
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Bone marrow;
RX MEDLINE=95308325; PubMed=7788527;
RA Nagase T., Miyajima N., Tanaka A., Suzuki T., Seki N., Sato S.,
RA Tabata S., Ishikawa K.-I., Kawabayashi Y., Kotani H., Nomura N.,
RT "Prediction of the coding sequences of unidentified human genes. III.
RT The coding sequences of 40 new genes (KIAA0081-KIAA0120) deduced by
RT analysis of cDNA clones from human cell line KG-1.";
RL DNA Res. 2:37-43(1995).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=lung;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Bhat N.K.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max A.M., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshyuki S., Carninci P., Prange C.,
RA Raha S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting J., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickens M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalley D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

```

[4]
RP SEQUENCE OF 4-391 FROM N.A. (ISOFORM 1), INTERACTION WITH WTL,
RP SUBCELLULAR LOCATION, AND TISSUE SPECIFICITY.
RC TISSUE=Fetal kidney;
RX MEDLINE=20458888; PubMed=11001926;
RA Little N.A., Hastie N.D., Davies R.C.;
RT 'Identification of WTAP, a novel Wilms' tumour 1-associating
RT protein.';
RL Hum. Mol. Genet. 9:2231-2239(2000).
RN [5]
RP SEQUENCE OF 112-396 FROM N.A. (ISOFORM 1).
RC TISSUE=Amnygdala;
RA Wambutt R., Heubner D., Meves H.-W., Weil B., Wiemann S.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SUBUNIT: Interacts with WTL.
CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q15007-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q15007-2; Sequence=VSP_010278, VSP_010279;
CC -!- TISSUE SPECIFICITY: Ubiquitous.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AF374416; AA54764.1; -
CC ENBL; D14661; BAA03495.1; -
CC ENBL; BC000383; AAH00383.1; -
CC ENBL; BC004432; AAH04432.1; -
CC ENBL; AJ276706; CAC10188.1; -
CC ENBL; AL583911; CAC29495.1; -
CC Genew; HGNC:16846; WTAP.
CC MIM; 605442; -
CC Alternative splicing; Nuclear protein.
CC VARSPLIC 151 151 S->R (in isoform 2).
CC VARSPLIC 152 396 /FTId=VSP_010278.
CC FT Missing (in isoform 2).
CC FT /FTId=VSP_010279.
CC SQ SEQUENCE 396 AA; 44243 MW; B4F61B3000F77A37 CRC64;
Query Match 99.1%; Score 1593; DB 1; Length 396;
Best Local Similarity 99.1%; Pred. No. 1.1e-72;
Matches 3.6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 MRLATKGQDIQECTIOYLVKQVQPSVAQLRSTWVDPAINFLLKWKGELEQTKDKLEQ 60
Db 78 MRLATKEQMEQECTIOYLVKQVQPSVAQLRSTWVDPAINFLLKWKGELEQTKDKLEQ 137
QY 61 AQNELSAWKFTPDSTQTKLMKACRMLIQENQELGRLSQGRIAGLEALQKYYSEEL 120
Db 138 AQNELSAWKFTPDSTQTKLMKACRMLIQENQELGRLSQGRIAGLEALQKYYSEEL 197
QY 121 KSSQDELNDFIIQLDDEVEGCMSTILVLOQQLKETRQQLAQYQQQSQASAPSTRTTAS 180
Db 198 KSSQDELNDFIIQLDDEVEGCMSTILVLOQQLKETRQQLAQYQQQSQASAPSTRTTAS 257
QY 181 EPEVQSEATSKDCSLRTNFPNGSSRRQRTSGSGFHRENTTDDFPSPGNGKSNSS 240
Db 258 EPEVQSEATSKDCSLRTNFPNGSSRRQRTSGSGFHRENTTDDFPSPGNGKSNSS 317
QY 241 EERTGRGSGYVQQLSAGVESVDSPTGENSELTHOSNDTDSHDPQEKAVSGKGNRTVG 300
Db 318 EERTGRGSGYVQQLSAGVESVDSPTGENSELTHOSNDTDSHDPQEKAVSGKGNRTVG 377
QY 301 SEHVQNGILDSSVNVQGSVL 319

```

```

DB 378 SRHVQGLDSSVNVQGSVL 396
|||||
RESULT 3
WTAP_MOUSE STANDARD; PRT; 396 AA.
AC Q9ER69;
DT 05-JUL-2004 (Rel. 44, Created)
DT 05-JUL-2004 (Rel. 44, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Wtaps' tumor 1-associating protein (Wtl-associating protein) (Putative
DE pre-mRNA splicing regulator female-lethal(2D) homolog).
GN Names=Wtap;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE OF 18-396 FROM N.A. (ISOFORM 1).
RP MEDLINE=20458888; PubMed=11001926;
RA Little N.A., Hastie N.D., Davies R.C.;
RT "Identification of WTAP, a novel Wilms' tumour 1-associating
RT protein."
RL Hum. Mol. Genet. 9:2231-2239(2000).
RN [2]
SEQUENCE FROM N.A. (ISOFORM 2).
RP STRAIN=C57BL/6J;
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt T.A., Brusci V., Chochia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Nunata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., King B.Z., Ringwald M.,
RA Sadelin A., Schneider C., Sempile C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Vitaranto R., Wagner L., Wahlstedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Waki J., Alzawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
RN [3]
SEQUENCE FROM N.A. (ISOFORM 2).
RC STRAIN=Czech II; Tissue=Breast tumor;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Berge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieff F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.D., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullah S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

```

```

RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- SUBMIT: Interacts with Wtl (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q9ER69-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9ER69-2; Sequence=VSP_010280, VSP_010281;
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ276707; CAC16790.1; -
DR EMBL; AK076111; BAC36191.1; -
DR EMBL; BC046416; AAH46416.1; -
DR MGD; MGI:1926395; Wtap.
KW Alternative splicing; Nuclear protein.
FT VARSPLIC 151 151 S -> R (in isoform 2).
FT VARSPLIC 152 396 Missing (in isoform 2).
FT VARSPLIC 152 396 Missing (in isoform 2).
FT VARSPLIC 152 396 Missing (in isoform 2).
SQ SEQUENCE 396 AA; 44215 MW; 63900F41F9D01777 CRC64;
Query Match 95.2%; Score 1531; DB 1; Length 396;
Best Local Similarity 94.7%; Pred. No. 1.5e-69;
Matches 302; Conservative 8; Mismatches 9; Indels 0; Gaps 0;
QY 1 MRLATKGODIIBCTTQYLYKQVQPSVAQLRSTMDPAINLFLKMKGELEQTKKLEQ 60
DB 78 MRLATKEQEMQECTTQYLYKQVQPSVAQLRSTMDPAINLFLKMKGELEQTKKLEQ 137
QY 61 AQNELSANKFPDSTQTKKLMKCRMLIQNQLGRQLSQGRIAQLEAEALOKKXSEEL 120
DB 138 AQNELSANKFPDSTQTKKLMKCRMLIQNQLGRQLSQGRIAQLEAEALOKKXSEEL 197
QY 131 KSSQDELNDFIQLDEVEGQSTILVLOQLKTRQQLAQYQQQSQASAPSTRTAS 180
DB 198 KSSQDELNDFIQLDEVEGQSTILVLOQLKTRQQLAQYQQQSQASAPSTRTSS 257
QY 181 EPVQAEVTSKDCRLANGPNSGSSRQRTSGSGFHREGSTPEDDPSSPGNKGKSNSS 240
DB 258 EPVQAEVTSKDCRLANGPNSGSSRQRTSGSGFHREGSTPEDDPSSPGNKGKSNSS 317
QY 241 ERTGRGSGGVNQLSAGYVDSPTGSENLSHQSNDDTSSHPQBEKAVSGKGRVTG 300
DB 318 ERTGRGSGGVNQLSAGYVDSPTGSENLSHQSNDDTSSHPQBEKAVSGKGRVTG 377
QY 301 SRHVQGLDSSVNVQGSVL 319
DB 378 SRHVQGLDSSVNVQGSVL 396
RESULT 4
Q6P4K5 PRELIMINARY; PRT; 393 AA.
ID Q6P4K5;
AC Q6P4K5;
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)

```

AAH63362;
25-MAR-2004 (TrEMBLrel. 27, Created)
DT 25-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 25-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein MGC75954.
GN MGC75954.
OS Xenopus tropicalis (Western clawed frog) [Silurana tropicalis].
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus; Silurana.
OX NCBI_TaxID=8364;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RC MEDLINE=22388257; PubMed=12477912;
RA Strausberg R.L., Feingold B.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T.I., Max S.I., Wang J., Hsieh P.,
RA Diatchenko L., Maruina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M.J., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Tohiyuki S., Carninci P., Prange C.,
RA Raha S.S., Lomuellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalls D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RC Klein S., Strausberg R.;
RA Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC063362; AAH63362.1; -
KW Hypothetical protein.
SQ SEQUENCE 393 AA; 44312 MW; 82983BAB02EA99AE CRC64;
Query Match 80.68; Score 1296.5; DB 2; Length 393;
Best Local Similarity 81.28; Pred. No. 9.2e-58;
Matches 260; Conservative 23; Mismatches 32; Indels 5; Gaps 2;
QY 1 MRLATGQDIQECTTQTQLVKVQQPSVAQLRSTWVDPAINLFFKMKGELOTKDKLEQ 60
Dd :::::
Dd 78 MRLATKEQEOMETTQTQLHLKVQQPSVAQLRATWVDPAINLFFKMAELEOTKDKLEQ 137
QY 61 AQNELSAWKTPPSQTGGKLMACRMILIOENELGRQLSQRIALAEIALAQKYSEEL 120
Dd :::::
Dd 138 AQNELSAWKTPPSQTGGKLMACRMILIOENELGRQLSQRIALAEIALAQKYSEEL 197
QY 121 KSSODELNDFIIQDDEVEGMOSTLVLCQKLKETRQOLAQYQQQSQAAPSRTTAS 180
Dd :::::
Dd 198 KSSODELNDFIIQDDEVEGMOSTLVLCQKLKDSRQLSOFQOO----TQSGNRTPSS 253
QY 181 EPVEQSATSKDCSRUTNGPNSGSSRRORTGS-GFHREGNTTEDDPSPGPNKGSSNS 239
Dd :::::
Dd 254 EKNDGETSGDKGRILINGPSNGGSHORTHSSVLGYREGSSTBEDFSPINEGLPNH 313
QY 240 SEERTGRGGGVYNQLSAGYESVDSPTGENSLTHQSNDDTSSHDPOEKAVSGKNRTV 299
Dd :::::
Dd 314 SEERTSEGSSVYNQLSTGESVDSPTGENSLTHQSNDDTSSHDPOEKAVSGKNRTV 373
QY 300 GSRHVQNGLDSSVNVQGSVL 319
Dd :::::
Dd 374 SSRHLQNLGDSSVNVQGSVL 393


```
RESULT 6
Q7SXL7 PRELIMINARY; PRT; 423 AA.
AC Q7SXL7;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Wlms, tumour 1-associating protein.
GN Name=wtap;
OS Brachydanio rerio (zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OC NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=whole body;
RX MEDLINE=22389257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heien F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalusz D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
RA and mouse cDNA sequences.";
RA proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=whole body;
RA Strausberg R.;
RA Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC055544; AAH55544.1;
SQ SEQUENCE 423 AA; 46151 MW; A2600E687DAA229C CRC64;

Query Match 71.3%; Score 1146.5; DB 2; Length 423;
Best Local Similarity 70.9%; Pred. No. 3.5e-50;
Matches 246; Conservative 25; Mismatches 47; Indels 29; Gaps 8;

QY 1 MRLATKGDIQECTTQIQLKQVQPSVAQLRSTWVDPAINLFLKMKGELEQTKLEQ 60
DB 78 MRLATKEQEMQECTTQIQLKQVQPSAAQLRSSWVDPAINLFLKMKAELEQTKLEQ 137
QY 61 AQNELSAWKFTPDSTQTKMLAKCMRLIQENQELGRQLSQRIQLAEALQKYSSEL 120
DB 138 AQNELSAWKFTPDSTQTKMLAKCMRLIQENQELGRQLSQRIQLAEALQKYSSEL 197
QY 121 KSSQDELNDFFIQLDEEVEGQSTILVLQQLKETROQLAQYQQQSQASAPTSRT--- 177
DB 198 KSSQDELNDFFIQLDEEVEGQSTILVLQQLKETROQLAQYQQQSQASAPTSRTSPS 257
QY 178 TASEPVESEATS-----KDCSLRTNGPNSGSSRQRTSGSGFHREGNTTDDFPSSP 230
DB 258 TASEPSTQSEPANASSNVKDCGRVNSGPNSSRGSGSSLYREASSADEYFPSP 317
QY 231 G-----NG-KKSSNSSSEERTG-RGSGGVNQLSAGYSDSPGTSNSLTHQSNDDTSS 282
DB 316 SVSSPTHDGKSLNSHSDAVSQRGEGYVITQLSAGYSDSPGTSNTHQSNDDTSDN 377
QY 283 HDPQEKAVSGKGNRTVGSRH-VQGLDSSVNV-----QGSVL 319
DB 378 ADSHEAAAAP-KGSRTAGTRHSTQNGLDSSAAVAATNTSNASAGSVL 423
```

```
RESULT 7
Q7KA73 PRELIMINARY; PRT; 412 AA.
AC Q7KA73;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE FL(2)D-s protein.
GN Name=fl(2)d;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20253098; PubMed=10790389;
RA Penalba L.O.F., Ruiz M.F., Ortega A., Granadino B., Vicente L.,
RA Segarra C., Valcarcel J., Sanchez L.;
RT "The Drosophila fl(2)d gene, required for female-specific splicing of
RT Sxl and tra pre-mRNAs, encodes a novel nuclear protein with a HQ-rich
RT domain.";
RL Genetics 155:129-139(2000).
DR EMBL; AJ243607; CAB46726.1;
SQ SEQUENCE 412 AA; 45055 MW; 17979281B158D581 CRC64;

Query Match 32.9%; Score 528.5; DB 2; Length 412;
Best Local Similarity 33.9%; Pred. No. 4.3e-19;
Matches 127; Conservative 62; Mismatches 113; Indels 73; Gaps 8;

QY 2 RLATKGDIQECTTQIQLKQVQPSVAQLRSTWVDPAINLFLKMKGELEQTKLEQ 61
DB 18 RLANKQEFDVVSQIAEYKAAQAPALALRALTALDPAVNLFFERLKKELKATKALEET 77
QY 62 QNELSAWKFTPDSTQTKMLAKCMRLIQENQELGRQLSQRIQLAEALQKYSSEL 121
DB 78 QNELSAWKFTPDSTQTKMLAKCMRLIQENQELGRQLSQRIQLAEALQKYSSEL 137
QY 122 SSQDELNDFFIQLDEEVEGQSTILVLQQLKETROQLAQYQQQSQ-----ASA 171
DB 138 KQSELDDFLQELDEDEVEGQSTILFLQQLKETTRDRIQTLEKNAQLKQKDEWAPA 197
QY 172 PSTRRTTASEPVESEATSQKDCSLRTNGPNSGSSRQRTSGSGFHREGNTTDDFPSS-- 229
DB 198 AATNGGT-NTTINKLETHEDAC-MANNPTNDC-----YNGNTNNEQIAAAPQ 244
QY 230 --PGNGKSSNSEERTGSGGVNQLSAGYSDSPGTSNSL----- 272
DB 245 IPPSDGSGNNGVAARLAKRNYQEEALPTVVVPTPTPGVNNVQEAAPPREVTA 304
QY 273 -----THQSNDDTSSHDPOB-----EKAVSGKGNRTVGSRHV----- 304
DB 305 PPKSKSLRGITTRNSQLSEDEHQVTPVAVPMVDNAVAGMASEEAAAAAANNNTG 364
QY 305 ---QGLDSSVNVQ 316
DB 365 IIPETGVQGVPEV 379

RESULT 8
Q9Y091 PRELIMINARY; PRT; 539 AA.
AC Q9Y091; Q9U974;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE FL(2)D protein.
GN Name=fl(2)d;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
```

OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RA SEQUENCE FROM N.A.
 RP Penelope L.O.F., Ruiz M.F., Granadino B., Vicente L., Ortega A.,
 RA Valcarlos J., Sanchez L.;
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ243599; CAB46637.1; ..
 DR FlyBase; FBgn0000662; fl(2)d.
 DR GO; GO:0005634; C:nucleus; IDA.
 DR GO; GO:0019099; P:female germ-line sex determination; NAS.
 DR GO; GO:0007539; P:primary sex determination, soma; NAS.
 DR GO; GO:0000381; P:regulation of alternative nuclear mRNA splicing; IPI.
 DR GO; GO:0007530; P:sex determination; IPI.
 DR InterPro; IPR009053; Pfam01d.
 SQ SEQUENCE 539 AA; 59608 MW; FB2F0579F83CFCD5 CRC64;
 Query Match 32.9%; Score 528.5; DB 2; Length 539;
 Best Local Similarity 33.9%; Pred. No. 5.8e-19;
 Matches 127; Conservative 62; Mismatches 113; Indels 73; Gaps 8;
 QY 2 RIATKGQDIQCTTQIQLKQVQPSVAQLRSTWVDPAINTLFFLMKGELEOTKDKLEQA 61
 Db 145 RLANKQEFQDVSVQIAEYKQAQPTALATLTDPAVNLFFELKKELKATKALEET 204
 QY 62 QNELSAWFTDSTQTKLMAKRMLOENQELGRLSQGRIAGRLQELAEALQKYSBELK 121
 Db 205 QNELSAWFTDSTQTKLMAKRMLOENQELGRLSQGRIAGRLQELAEALQKYSBELK 121
 QY 122 SSQBELNFIQLDVEEVMGSTILVLQQLKQETROQLAQVQQQSQ-----ASA 171
 Db 265 KSQSELDFLQELDEVEGMSTILFLQQLKQETROQLAQVQQQSQ-----ASA 171
 QY 172 PSTERTASEPVEQSEATSKDCSLTNGPSNGSSORTSGSGFHRGENTEDDFPSS-- 229
 Db 325 AATNGGT-NTTINKLETHEDAC-MANNPTNPD-----YNGNTNNEQIAAAPQ 371
 QY 230 -PGNGKSSNSSEERTGGSGVYNQLSAGYSDVDSPTGSENLS----- 272
 Db 372 IPPDGGSGNMGNAARLARKNYEBALEALPTVVVPTPVGNNVQAPPIREVTAPRTL 431
 QY 273 -----THOSNPTDSSHPQE-----EKAVSGKGNRTVGSRHV----- 304
 Db 432 PPKSKLGRITRANSQLEEDHQVPTTPVAVPMVDVNAVAGMASEEAAAAAANNNTG 491
 QY 305 ---QNGLDSSVNVVG 316
 Db 492 IIPETGVGVFVEG 506
 RESULT 9
 Q8ML74
 ID Q8ML74 PRELIMINARY; PRT; 412 AA.
 AC Q8ML74
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE CG6315-PB
 GN Name=fl(2)d; ORFNames=CG6315;
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RA SEQUENCE FROM N.A.
 RP MEDLINE=20196006; PubMed=107311132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
 RA Lewis S.E.;
 RL "Annotation of the Drosophila melanogaster euchromatic genome: a systematic review.";
 RT Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
 RN [5]
 RA SEQUENCE FROM N.A.
 RP FLYBASE;
 RN Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RL [6]
 RA SEQUENCE FROM N.A.
 RP FLYBASE;

RA Abril J.P., Achayani A., An H.J., Andrews-Pfannkuch C., Baldwin D.,
 RA Balaw R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlike C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.B., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinart K., Remington K., Saunders R.D., Scheefer F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Turner R., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zavari J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN [12]
 RA SEQUENCE FROM N.A.
 RP MEDLINE=22426065; PubMed=12537568;
 RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
 RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
 RA Pacle J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
 RA Svirskas R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
 RA Weinstein G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
 RT "Finishing a whole-genome shotgun: release 3 of the Drosophila melanogaster euchromatic genome sequence.";
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
 RN [13]
 RA SEQUENCE FROM N.A.
 RP MEDLINE=22426070; PubMed=12537573;
 RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
 RA Ashburner M., Celniker S.E.;
 RT "The transposable elements of the Drosophila melanogaster euchromatic genome: a genomics perspective.";
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
 RN [14]
 RA SEQUENCE FROM N.A.
 RP MEDLINE=22426069; PubMed=12537572;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
 RA Smith C.D., Tupy J.L., Whittif E.J., Bayraktaroglu L., Berman B.P.,
 RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a systematic review.";
 RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
 RN [15]
 RA SEQUENCE FROM N.A.
 RP FLYBASE;
 RN Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RL [6]
 RA SEQUENCE FROM N.A.
 RP FLYBASE;

```

RA Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
RA EMBL; AE003817; AA071007.1; -.
RA FlyBase; FBgn0000662; fl12/d.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0019099; P:female germ-line sex determination; NAS.
DR GO; GO:0007539; P:pregnancy sex determination, soma; NAS.
DR GO; GO:0000381; P:regulation of alternative nuclear mRNA splicing; IPI.
DR GO; GO:0007530; P:sex determination; IPI.
DR InterPro; IPR009053; Pfam domain.
SQ SEQUENCE 412 AA; 45099 MW; EC8782C615653DD2 CRC64;

Query Match 32.7%; Score 525.5; DB 2; Length 412;
Best Local Similarity 33.9%; Pred. No. 6e-19;
Matches 127; Conservative 61; Mismatches 114; Indels 73; Gaps 8;

QY 2 RLATKGODIQECTQIQLKQVOQPSVAQLRSTWDPAINLPFLKVKGLQETKDKLEOA 61
Db 18 RLANKGEQFYQSYIAEYKAQAQAPTALALRTALDPAVLLPFLKELKATKALEET 77
QY 62 QNELSAKFKTPDSOTGKMLMAKCRMLIQENQELGRLSQGLRQIAQLPAELALOKKYSEELK 121
Db 78 QNELSAKFKTPDSNTGKRLMAKCRLLYQENELGKMTSNGRLAKLETALAMQKSFSEEVK 137
QY 122 SSQBELNDPIQLDDEVEGHQSTILVLOQQLKTRQOLAQYQQQSQ-----ASA 171
Db 138 KSQSELDLFLQELDEVEGQSTILFLQQLKTRDRIQTLKXNAQLKQAIKDEVVAPA 197
QY 172 PSTERTTASPEVSEOSTAKDCSLTNGPSNGSSRSQRTSGGFHREGNTTDDFPSSP- 230
Db 198 AATNGGT-NTINKLEIHDAC-MANNPNDP-----YNGNTNNQIAAVPQ 244
QY 231 ---GNGKNSNSBERTGRGSGGVNQLSAGYSVDSPGTSNSL----- 272
Db 245 IPLSDGSGNMGNAARLARKNYQEBALEPTVVVPTPTPGVNNVQBPPIREVTAPRTL 304
QY 273 -----THOSNTDSSHDPQE-----EKAVSGKNRTVGSRHV----- 304
Db 305 PPKSKLGRITRRNSQLEDHQPTTPVAVPMVDNAVAGMASEEAAAAAANNNTG 364
QY 305 ---QNGLDSSVNVQ 316
Db 365 IIPETGVGVGVPEVG 379

RESULT 10
QYV6T8 PRELIMINARY; PRT; 536 AA.
AC QYV6T8;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
DE CG6315-PA (LD21616p).
GN Name=fl(2)d; ORFNames=CG6315;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Prorygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adam M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazell R.G., Chape M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.D.,
RA April J.F., Agbayani A., An H.J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

```

```

de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,
Foeller C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
Reinert K., Remington K., Saunders R.D., Scheefel F., Shen H.,
Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
Svirskas R., Tecor C., Turner R., Venter E., Wang A.H., Wang X.,
Wang Z.Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
Williams S.M., Woodger, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zhou L.,
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
"The genome sequence of Drosophila melanogaster."
Science 287:2185-2195(2000).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champagne M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G., Venter C.,
RA Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
"Finishing a whole-genome shotgun: release 3 of the Drosophila
RA melanogaster euchromatic genome sequence."
Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
[3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
"the transposable elements of the Drosophila melanogaster euchromatin:
a genomics perspective."
Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
[4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Bayraktaroglu L., Berman B.P.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Drysdale R.A.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
"Annotation of the Drosophila melanogaster euchromatic genome: a
systematic review."
Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
[5]
RP SEQUENCE FROM N.A.
RX FLYBASE;
RN Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
[6]
RP SEQUENCE FROM N.A.
RX FLYBASE;
RN Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
[7]
RP SEQUENCE FROM N.A.
RX STRAIN=Berkeley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,

```


OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid1; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=99024949; PubMed=9807817;
RX Faure J.-D., Gingerich D., Howell S.H.;
RT "An Arabidopsis immunophilin, AtFKBP12, binds to AtPIP37 (FKBP
interacting protein) in an interaction that is disrupted by FK506.";
RL Plant J. 15:783-789 (1998).
RN [2]
RN SEQUENCE FROM N.A.
RP Faure J.-D., Gingerich D., Howell S.H.;
RX Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RN SEQUENCE FROM N.A.
RP Bloembergen H., Mewes H.W., Lemcke K., Mayer K.F.X., Quetier F.,
RA Salancubut M.;
RX Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RN SEQUENCE FROM N.A.
RP EU Arabidopsis sequencing project;
RX Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RN SEQUENCE FROM N.A.
RP Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,
RX Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
RA Seki M., Shinn P., Southwick A., Shinzaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RX Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
RN [6]
RN SEQUENCE FROM N.A.
RP Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,
RA Deng J.M., Goldsmith A.D., Lee J.M., Oncidera C.S., Quach H.L.,
RX Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
RA Seki M., Shinn P., Southwick A., Shinzaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RX Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
RN [7]
RN SEQUENCE FROM N.A.
RP EMBL; AF084570; AAC72922.1; -;
DR EMBL; AL132957; CAB70991.1; -;
DR EMBL; AY063940; AAL36296.1; -;
DR EMBL; AY113943; AAM44991.1; -;
DR PIR; T47576; T47576.
SQ SEQUENCE 330 AA; 37214 MW; 6B4070DSAD332004 CRC64;
Query Match 17.5%; Score 282; DB 2; Length 330;
Best Local Similarity 33.5%; Pred. No. 8.4e-07;
Matches 55; Conservative 51; Mismatches 58; Indels 0; Gaps 0;
QY 1 MRLATKGDIOECTTQIQLKQVQSPVAQLRSTWDPAINLFLKMKGELEOTKLEQ 60
DB 136 VQAKRQEMAEKSAVRLDKSLQKLPASMQARRLLDPAIHEEFTLNKLVVEKKVKE 195
QY 61 AQNELSAWKFTPDSTQTKMLMAKCRMLIQENQELGRQLSQRIAGLEALQKYSLEL 120
DB 196 LQDNIAAVTFTPQSKNGKMLMAKCRMLTQEEENEIGHQAAGKIHELAIKLAMQKSQNAEL 255
QY 121 KSSQDELNDFTIQLDEEVEGQSTILVLOQLKKTROQLAQYQ 164
DB 256 RSQFEGLYKHMEELNDVRSNETVILQLEKKEIEIRVKK 299
RESULT 14
Q6PNZ9 PRELIMINARY; PRT; 267 AA.
ID Q6PNZ9
AC Q6PNZ9;

DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE FKBP12-like protein (Fragment).
OS Mirabilis jalapa (Garden four-o'clock).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllales; Nyctaginaceae; Mirabilis.
OX NCBI_TaxID=3538;
RN [1]
RN SEQUENCE FROM N.A.
RP TISSUE=Petiole abscission zone;
RA Meir S., Hunter D.A., Chen J.-C., Reid M.S.;
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY589703; AAT07458.1; -;
FT NON_TER 1
FT NON_TER 267
SQ SEQUENCE 267 AA; 30178 MW; D2A198B6C3A3A4AE CRC64;
Query Match 13.6%; Score 218; DB 2; Length 267;
Best Local Similarity 37.3%; Pred. No. 0.0011;
Matches 44; Conservative 32; Mismatches 42; Indels 0; Gaps 0;
QY 4 ATKGDIOECTTQIQLKQVQSPVAQLRSTWDPAINLFLKMKGELEOTKLEQAOQN 63
DB 150 AKRQEIATLKAARELSQLKPLSMQTRLLDPAIHEEFTLNKLVVEKKVKELOD 209
QY 64 ELSAWKFTPDSTQTKMLMAKCRMLIQENQELGRQLSQRIAGLEALQKYSLEL 121
DB 210 NINAVNFTANSKMGKMLMAKCKTQEEENEIGHQAAGKIHELAIKLAMQKSQNAEL 267
RESULT 15
AAT07458 PRELIMINARY; PRT; 267 AA.
ID AAT07458
DT 12-MAY-2004 (TrEMBLrel. 27, Created)
DT 12-MAY-2004 (TrEMBLrel. 27, Last sequence update)
DT 12-MAY-2004 (TrEMBLrel. 27, Last annotation update)
DE FKBP12-like protein (Fragment).
OS Mirabilis jalapa (Garden four-o'clock).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllales; Nyctaginaceae; Mirabilis.
OX NCBI_TaxID=3538;
RN [1]
RN SEQUENCE FROM N.A.
RP TISSUE=Petiole abscission zone;
RA Meir S., Hunter D.A., Chen J.-C., Reid M.S.;
RL "Gene changes associated with the acquisition of abscission zone
competence in Mirabilis jalapa.";
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY589703; AAT07458.1; -;
FT NON_TER 1
FT NON_TER 267
SQ SEQUENCE 267 AA; 30178 MW; D2A198B6C3A3A4AE CRC64;
Query Match 13.6%; Score 218; DB 2; Length 267;
Best Local Similarity 37.3%; Pred. No. 0.0011;
Matches 44; Conservative 32; Mismatches 42; Indels 0; Gaps 0;
QY 4 ATKGDIOECTTQIQLKQVQSPVAQLRSTWDPAINLFLKMKGELEOTKLEQAOQN 63
DB 150 AKRQEIATLKAARELSQLKPLSMQTRLLDPAIHEEFTLNKLVVEKKVKELOD 209
QY 64 ELSAWKFTPDSTQTKMLMAKCRMLIQENQELGRQLSQRIAGLEALQKYSLEL 121
DB 210 NINAVNFTANSKMGKMLMAKCKTQEEENEIGHQAAGKIHELAIKLAMQKSQNAEL 267
Search completed: November 14, 2004, 12:36:30
Job time : 194 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2004, 12:14:48 ; Search time 156 Seconds
(without alignments)
733.555 Million cell updates/sec

Title: US-10-030-389-2
Perfect score: 1608
Sequence: 1 MRLATKGQDIQECTTQIQL.....GSRHVQNGLDSSVNVQGSVL 319

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseqp23Sep04:.*
1: geneseqp1980s:.*
2: geneseqp1990s:.*
3: geneseqp2000s:.*
4: geneseqp2001s:.*
5: geneseqp2002s:.*
6: geneseqp2003as:.*
7: geneseqp2003bs:.*
8: geneseqp2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1608	100.0	319	4	AAB67048 Human imm
2	1570	97.6	388	6	ABO53045 Human put
3	525.5	32.7	536	4	ABB62262 Drosophil
4	405.	25.2	671	4	ABG20586 Novel hum
5	357	22.2	151	6	ABO53092 Human put
6	357	22.2	151	8	ADO58680 Human reg
7	282	17.5	278	3	AG41464 Arabidops
8	282	17.5	330	3	AG41463 Arabidops
9	271	16.9	186	3	AG41465 Arabidops
10	256	15.9	352	8	ADM48175 Polypepti
11	178	11.1	102	4	ABG10314 Novel hum
12	166	10.3	524	6	ABU29734 Protein e
13	166	10.3	525	7	AC95468 E. faeciu
14	156.5	9.7	511	5	ABG93255 C. albica
15	156	9.7	897	2	AAW2751 Murine EG
16	156	9.7	897	2	AAW47118 Murine ep
17	156	9.7	897	2	AAW94406 Murine ep
18	152.5	9.5	1031	4	AAU35339 Enterococ
19	150	9.3	431	7	ADD93650 Streptoco
20	150	9.3	456	5	ABE5584 Lactococ
21	149	9.3	431	7	ADD93653 Streptoco
22	147	9.1	912	4	ABE60403 Drosophil
23	147	9.1	2112	4	ABE60403 Drosophil
24	146	9.1	431	7	ADD93649 Streptoco
25	145	9.0	3502	4	ABB58382 Drosophil

26	144.5	9.0	877	8	ADN72429 Thale cre
27	144	9.0	461	2	AAR14530 USP45 pro
28	144	9.0	461	2	AAR14530 MSP encod
29	144	9.0	716	2	AY22176 Drosophil
30	144	9.0	1807	8	ADO17160 Novel exp
31	142.5	8.9	744	4	ABU71321 Drosophil
32	142.5	8.9	1061	6	ABU29451 Protein e
33	141.5	8.8	432	7	ADD93651 Streptoco
34	141.5	8.8	562	4	ABB60336 Drosophil
35	141	8.8	294	5	ABP41663 Human ova
36	141	8.8	1802	3	AA783170 Cell wall
37	141	8.8	1802	3	AA783170 Staph. ep
38	141	8.8	3248	2	AA99795 Kinetoch
39	140	8.7	108	4	ABG20587 Novel hum
40	140	8.7	1031	7	ADB67093 Kinesin h
41	140	8.7	1032	8	ADL99364 Nancstruc
42	140	8.7	1091	4	ABB68988 Drosophil
43	140	8.7	1920	5	ABG97508 Human NOV
44	139.5	8.7	1135	3	AA768784 Amino aci
45	139	8.6	1017	7	ADJ68352 Human hea

ALIGNMENTS

RESULT 1
AAB67048
ID AAB67048 standard; protein; 319 AA.
XX
AC AAB67048;
XX
DT 09-APR-2001 (first entry)
XX
DE Human immune response molecule (IMUN) protein SEQ ID NO: 2.
XX
KW Human; IMUN; immune response molecule; autoimmune disorder;
KW inflammatory disorder; cell proliferation disorder; cancer.
XX
OS Homo sapiens.
XX
PN WO200102569-A2.
XX
PD 11-JAN-2001.
XX
PF 06-JUL-2000; 2000WO-US018505.
XX
PR 06-JUL-1999; 99US-0142572P.
PR 09-SEP-1999; 99US-0153170P.
XX
(INCY-) INCYTE GENOMICS INC.
PA Tang YT, Yue H, Yang J, Azimzai Y, Baughn MR, Lu DAM;
PI
XX
DR WPI; 2001-123113/13.
DR N-PSDB; AAF31245.
XX
PT Novel human immune response molecules (IMUN), useful for diagnosing,
PT treating and preventing disorders associated with abnormal expression of
PT IMUN, e.g. Addison's disease, allergies, anemia, asthma and
XX atherosclerosis.
PS Claim 2; Page 82; 95pp; English.
XX
CC The present invention provides the protein and coding sequences for ten
CC human immune response molecules (designated IMUN-1 to IMUN-10). The
CC sequences can be used in the diagnosis, treatment and prevention of
CC autoimmune and inflammatory disorders such as AIDS, atherosclerosis,
CC asthma, allergies, Crohn's disease, multiple sclerosis, irritable bowel
CC syndrome, psoriasis, rheumatoid arthritis and infections, and cell
XX proliferation disorders including arteriosclerosis, cirrhosis and cancer
SQ Sequence 319 AA;

Query Match 100.0%; Score 1608; DB 4; Length 319;
 Best Local Similarity 100.0%; Pred. No. 5.7e-121; Indels 0; Gaps 0;
 Matches 319; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRLATKGQDIQECTTQIQYLVKQVQPSVAQLRSTWVDPAINLFFFLKMGELQTKDKLEQ 60
 DB 1 MRLATKGQDIQECTTQIQYLVKQVQPSVAQLRSTWVDPAINLFFFLKMGELQTKDKLEQ 60

QY 61 AQNELSAWKFTPDSTGKMLAKCRLMIQENQELGRQLSQGRIALQLEALQKYSSEL 120
 DB 61 AQNELSAWKFTPDSTGKMLAKCRLMIQENQELGRQLSQGRIALQLEALQKYSSEL 120

QY 121 KSSQDELNDFFIQLDEEVEGQSTILVLQQLKTRQQLAQYQQQSQASAPSTRITAS 180
 DB 121 KSSQDELNDFFIQLDEEVEGQSTILVLQQLKTRQQLAQYQQQSQASAPSTRITAS 180

QY 181 EPVEQSEATSKDCSRLTNGPNSGSSRRQRTSGSGFHREGNTTDDFPSSPGNGKSNSS 240
 DB 181 EPVEQSEATSKDCSRLTNGPNSGSSRRQRTSGSGFHREGNTTDDFPSSPGNGKSNSS 240

QY 241 ERTGRGGSGYVNLQSLAGYVESVDSPGSENLSLTHQSNDTSSHDPPQBEKAVSGKGNRTVG 300
 DB 241 ERTGRGGSGYVNLQSLAGYVESVDSPGSENLSLTHQSNDTSSHDPPQBEKAVSGKGNRTVG 300

QY 301 SRHVQGLDSSVNVQGSVL 319
 DB 301 SRHVQGLDSSVNVQGSVL 319

RESULT 2
 ABO53045
 ID ABO53045 standard; protein; 388 AA.

XX AC ABO53045;
 XX DT 10-OCT-2003 (first entry)
 XX DE Human putative spliceosome associated protein (SAP) #21.
 XX KW Human; SAP; spliceosome associated protein; ribonucleoprotein;
 KW RNP complex; RNA affinity substrate; RNP assembly sequence;
 KW spliceosomal complex; hnRNP complex; mRNA export complex;
 KW mRNA localisation complex; RNA editing complex; intron complex;
 KW H complex; telomerase complex; fragile X protein complex;
 KW reverse transcriptase complex; gene splicing complex.
 XX OS Homo sapiens.
 XX US2003068803-A1.
 XX PD 10-APR-2003.
 XX PF 14-JAN-2002; 2002US-00047991.
 XX PR 12-JAN-2001; 2001US-0261521P.
 XX PA (REED/) REED R.
 XX PA (ZHOU/) ZHOU Z.
 XX PI Reed R, Zhou Z;
 XX DR WPI; 2003-540885/51.
 XX PT Isolating ribonucleoprotein complex, by contacting RNA affinity substrate
 PT having ribonucleoprotein assembly sequence and affinity tag, with protein
 PT mixture, subjecting complex formed to chromatography, affinity selection.
 XX PS Claim 24; Page; 39pp; English.
 XX CC The invention relates to forming (M1) an isolated ribonucleoprotein (RNP)
 CC complex (C), involves contacting an RNA affinity substrate (S) comprising
 CC an RNP assembly sequence (AS) and an affinity tag, with a protein mixture
 CC to permit formation of (C) on AS, subjecting (C) to chromatographic

CC separation, and subjecting (C) to affinity selection, where the affinity
 CC tag (e.g. bacteriophage MS2 coat protein in a fusion protein with E. coli
 CC maltose binding protein) binds to an affinity matrix. Also included are
 CC an isolated spliceosome preparation (isolated by (M1)), a RNA comprising
 CC an RNP complex binding site and at least one phase coat protein
 CC recognition site, a nucleic acid encoding the RNA, and treating (M2) a
 CC subject having a disorder associated with abnormal RNP complexes (by
 CC obtaining a sample of cells from a subject, purifying RNP complexes from
 CC the cells of the subject by (M1), determining the presence in the
 CC purified RNP complexes of one or more proteins, and normalising the
 CC amount of RNPs in the subject. (M1) is useful for forming an isolated RNP
 CC complex selected from a spliceosomal complex (selected from E, A, B and C
 CC complex), an hnRNP complex, an mRNA export complex, an mRNA localisation
 CC complex, an RNA editing complex, an intron complex, or an H complex. (M1)
 CC is useful in a diagnostic assay for determining whether a subject has
 CC abnormal RNP complexes, (M2) is useful for treating a subject having a
 CC disorder associated with abnormal RNP complexes. (M1) is useful for
 CC forming an isolated RNP complex such as a telomerase complex, a fragile X
 CC protein complex, a reverse transcriptase complex or a gene splicing
 CC complex. The present sequence represents a putative novel human
 CC spliceosome associated protein (SAP) isolated by the methods of the
 CC invention. Note: The present sequence is not shown in the specification
 CC but was obtained from Genbank or Swissprot using the information provided
 CC in table 2 of the specification
 XX SQ Sequence 388 AA;

Query Match 97.6%; Score 1570; DB 6; Length 388;
 Best Local Similarity 99.0%; Pred. No. 8.6e-118;
 Matches 311; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MRLATKGQDIQECTTQIQYLVKQVQPSVAQLRSTWVDPAINLFFFLKMGELQTKDKLEQ 60
 DB 75 MRLATKEQEMQECTTQIQYLVKQVQPSVAQLRSTWVDPAINLFFFLKMGELQTKDKLEQ 134

QY 61 AQNELSAWKFTPDSTGKMLAKCRLMIQENQELGRQLSQGRIALQLEALQKYSSEL 120
 DB 135 AQNELSAWKFTPDSTGKMLAKCRLMIQENQELGRQLSQGRIALQLEALQKYSSEL 194

QY 121 KSSQDELNDFFIQLDEEVEGQSTILVLQQLKTRQQLAQYQQQSQASAPSTRITAS 180
 DB 195 KSSQDELNDFFIQLDEEVEGQSTILVLQQLKTRQQLAQYQQQSQASAPSTRITAS 254

QY 181 EPVEQSEATSKDCSRLTNGPNSGSSRRQRTSGSGFHREGNTTDDFPSSPGNGKSNSS 240
 DB 255 EPVEQSEATSKDCSRLTNGPNSGSSRRQRTSGSGFHREGNTTDDFPSSPGNGKSNSS 314

QY 241 ERTGRGGSGYVNLQSLAGYVESVDSPGSENLSLTHQSNDTSSHDPPQBEKAVSGKGNRTVG 300
 DB 315 ERTGRGGSGYVNLQSLAGYVESVDSPGSENLSLTHQSNDTSSHDPPQBEKAVSGKGNRTVG 374

QY 301 SRHVQGLDSSVNV 314
 DB 375 SRHVQGLDSSVNV 388

RESULT 3
 ABB62262
 ID ABB62262 standard; protein; 536 AA.
 XX AC ABB62262;
 XX DT 26-MAR-2002 (first entry)
 XX DE Drosophila melanogaster polypeptide SEQ ID NO 13578.
 XX KW Drosophila; developmental biology; cell signalling; insecticide;
 XX KW pharmaceutical.
 XX OS Drosophila melanogaster.
 XX PN WO200171042-A2.
 XX XX

PD 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US009231.
XX 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX (PEKE) PE CORP NY.
PA Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX N-PSDB; ABL06365.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX Disclosure; SEQ ID NO 13578; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175), and the encoded proteins (AB857737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 536 AA;
XX Query Match 32.7%; Score 525.5; DB 4; Length 536;
XX Best Local Similarity 33.9%; Pred. No. 1.8e-33;
XX Matches 127; Conservative 61; Mismatches 114; Indels 73; Gaps 8;
QY 2 RLATKGDIORCTTQIQVLKQVQPSVAQLRSTWVDPAINLFFLMKGELEQTKDKLEQA 61
DB 142 FLANKEQFQDYVSGIAYKQAQAPTALATLTDPAVNLFFERKKELKATKALEET 201
QY 62 QNELSAWFTPDSTGKMLAKMRLIOENQELGRLSQGRIAGLEALALOKKYSLELK 121
DB 202 QNELSAWFTPDSTGKMLAKRLLYOENELGKMTSNGRLAKLETALAMQKSFSEVK 261
QY 122 SSODELNFILQDDEVEGMOSTILVLQQLKETROQLAQYQQOQSQ-----ASA 171
DB 252 KSQSELDLFLQELDDVEGMOSTILFLQQLKETROQLAQYQQOQSQ-----ASA 321
QY 172 PSTRTTASPEVQSEATSKDCSRLTNGPSNGSSSRQRTSGSGFHRGNTTDEDFPSP- 230
DB 322 AATNGGT-NTTINKLETHEDAC-MANNPTNPD-----YNGNTNNEQIAAVPQ 368
QY 231 ---GNGKSSNSSEBRTGSGGYVQLSAGVESVDSPTGSENIL----- 272
DB 369 IPLSDDGNGMGAARLARKYQEEALPTVVVPTPTVGNVQEAAPPREVTAPRTL 428
QY 273 -----THQNDTSDSHDPOE-----EKAVSGKGNRTVGSRHV----- 304
DB 429 PPKSKLGRITFRNSQLEEDHQPTTTPVAVPMIVDNAVAGVASEAAAAAANNVNTG 488
QY 305 ---QNGLDSSVNVQ 316
DB 489 IIPETGVQGVPEV 503
RESULT 4
ABG20586
ID ABG20586 standard; protein; 671 AA.
XX
AC ABG20586;
XX
DT 13-FEB-2002 (first entry)

XX Novel human diagnostic protein #20577.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US008631.
XX 31-MAR-2000; 2000US-00540217.
XX 23-AUG-2000; 2000US-00649157.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX N-PSDB; AAS64773.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX Claim 20; SEQ ID NO 50945; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX Sequence 671 AA;
XX Query Match 25.2%; Score 405; DB 4; Length 671;
XX Best Local Similarity 53.2%; Pred. No. 1.2e-23;
XX Matches 98; Conservative 2; Mismatches 23; Indels 32; Gaps 5;
QY 126 ELNDFIQLDDEVEGQSTILVLQQLKETROQLAQYQQOQSQASAPSTRRTTASEPVEQ 195
DB 381 ELNDFIQLDDEVEGQSTILVLQQLKETROQLAQYQQOQSQASAPSTRRTTASEPVEQ 440
QY 186 SEATSKDCSRLTNGPSNGSS-----SRQRTSGSGFHRG-----NTTEDD 225
DB 441 SEATSKDCSRLTNGPSNGSSPFRRGRSHSPAAPRGRGEGREAVGLARSGRGTMEQL 500
QY 226 FPSG-----PG-----NGMKSSNS--EERTGRGG 248
DB 501 LPGNTPPFPFGAWRELGRGPSSASSVREAAEGSGG 535
RESULT 5

QY	1	MRLATKQDIQECTTQIQYLKQVQPSVAQLRSTWVDPAINFLLKMKGELEQTKDKLEQ	60
DB	78	MRLATKEQEMQECTTQIQYLKQVQPSVAQLRSTWVDPAINFLLKMKGELEQTKDKLEQ <td>137</td>	137
QY	61	AQNELSAWKFTPD 73	
DB	138	AQNELSAWKFTPD 150	
RESULT	6		
ID	ADO58680	standard; protein; 151 AA.	
XX	ADO58680		
AC	ADO58680		
DT	15-JUL-2004	(first entry)	
XX	Human regulatory molecule HRM-1.		
XX	cytostatic; immunomodulator; agonist; antagonist; gene therapy;		
KW	human regulatory molecule; HRM; disease development; cell proliferation;		
KW	immune response; cancer.		
XX	Homo sapiens.		
OS	US2002058264-A1.		
PN	16-MAY-2002.		
XX	26-SEP-2001; 2001US-00840787.		
PD	23-SEP-1997; 97US-00933750.		
XX	20-JAN-1999; 99US-00234613.		
PR	03-MAR-2000; 2000US-00518865.		
XX	(INCY-) INCYTE PHARM INC.		
PA	Lal P, Hillman JL, Bandman O, Shah P, Au-Young J, Yue H;		
XX	Guegler KJ, Corley NC;		
PI	WPI; 2004-459763/43.		
DR	N-PSDB; ADO58729.		
XX	New human regulatory molecules, useful in the diagnosis and treatment of		
XX	cancer and immune disorders.		
PS	Claim 1; SEQ ID NO 1; 116pp; English.		
CC	The invention describes human regulatory molecules (HRM) (I) selected		
CC	from a group comprising the fully defined amino acid sequences of SEQ ID		
CC	NOs: 1-49. Also described are: an isolated polynucleotide (II) comprising		
CC	a nucleic acid sequence encoding (I) or the complement of the		
CC	polynucleotide (SEQ ID NOS:50-98); a composition comprising (II) and a		
CC	reporter molecule; an expression vector containing (II); a host cell		
CC	containing the vector; detecting (M1) expression of a nucleic acid in a		
CC	sample; screening (M2) a plurality of molecules to identify a ligand; a		
CC	diagnosing (M3) a disease associated with gene expression in a sample		
CC	containing nucleic acids; a composition comprising (I) and a		
CC	pharmaceutical carrier or a labeling moiety; screening (M4) a plurality		
CC	of molecules to identify a ligand; preparation and purification of		
CC	antibodies; an antibody which specifically binds to (I); and detecting		
CC	protein expression in a sample. The new human regulatory protein		
CC	molecules which are expressed during disease development and the		
CC	polynucleotides which encode them satisfies a need in the art by		
CC	providing compositions which are useful in the diagnosis and treatment of		
CC	diseases associated with cell proliferation, particularly immune		
CC	responses and cancers. This is the amino acid sequence of a human		
CC	regulatory molecule.		
XX	Sequence 151 AA;		
SQ	Query Match	22.2%; Score 357; DB 8; Length 151;	

Tue Nov 16 05:47:32 2004

us-10-030-389-2.rag

Best Local Similarity 95.9%; Pred. No. 1.le-20;		Matches 70; Conservative 2; Mismatches 1; Indels 0; Gaps 0;	
QY	1 MRLATKQDIOECTTOIQYLKVOQPSVAGLRSTWVDPAINFLLKWKGELEQTKDKLQ 60		
Db	78 MRLATKEQNEQETTQIQLKVOQPSVAGLRSTWVDPAINFLLKWKGELEQTKDKLQ 137		
QY	61 AQNELSAWKTPD 73		
Db	138 AQNELSAWKTPD 150		
RESULT 7			
ID	AAQ41464		
AC	AAQ41464		
XX	18-OCT-2000 (first entry)		
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 51592.		
KW	Protein identification; signal transduction pathway; metabolic pathway;		
KW	hybridisation assay; genetic mapping; gene expression control; promoter;		
KW	termination sequence.		
OS	Arabidopsis thaliana.		
FN	EP1033405-A2.		
PD	06-SEP-2000.		
PF	25-FEB-2000; 2000EP-00301439.		
XX	25-FEB-1999; 99US-0121825P.		
PR	05-MAR-1999; 99US-0123180P.		
PR	09-MAR-1999; 99US-0123548P.		
PR	23-MAR-1999; 99US-0125788P.		
PR	25-MAR-1999; 99US-0126264P.		
PR	29-MAR-1999; 99US-0126785P.		
PR	01-APR-1999; 99US-0127462P.		
PR	06-APR-1999; 99US-0128234P.		
PR	08-APR-1999; 99US-0128714P.		
PR	16-APR-1999; 99US-0129845P.		
PR	19-APR-1999; 99US-0130077P.		
PR	21-APR-1999; 99US-0130449P.		
PR	23-APR-1999; 99US-0130510P.		
PR	23-APR-1999; 99US-0130891P.		
PR	28-APR-1999; 99US-0131449P.		
PR	30-APR-1999; 99US-0132048P.		
PR	04-MAY-1999; 99US-0132407P.		
PR	04-MAY-1999; 99US-0132484P.		
PR	05-MAY-1999; 99US-0132485P.		
PR	06-MAY-1999; 99US-0132486P.		
PR	06-MAY-1999; 99US-0132487P.		
PR	07-MAY-1999; 99US-0132863P.		
PR	11-MAY-1999; 99US-0134256P.		
PR	14-MAY-1999; 99US-0134218P.		
PR	14-MAY-1999; 99US-0134219P.		
PR	14-MAY-1999; 99US-0134221P.		
PR	14-MAY-1999; 99US-0134370P.		
PR	18-MAY-1999; 99US-0134768P.		
PR	19-MAY-1999; 99US-0134941P.		
PR	20-MAY-1999; 99US-0135124P.		
PR	21-MAY-1999; 99US-0135353P.		
PR	24-MAY-1999; 99US-0135629P.		
PR	25-MAY-1999; 99US-0136021P.		
PR	27-MAY-1999; 99US-0136392P.		
PR	28-MAY-1999; 99US-0136782P.		
PR	01-JUN-1999; 99US-0137222P.		
PR	03-JUN-1999; 99US-0137528P.		
PR	04-JUN-1999; 99US-0137503P.		
PR	07-JUN-1999; 99US-0137724P.		
PR	08-JUN-1999; 99US-0138094P.		
PR	10-JUN-1999; 99US-0138540P.		
PR	10-JUN-1999; 99US-0138847P.		
PR	14-JUN-1999; 99US-0139119P.		
PR	16-JUN-1999; 99US-0139452P.		
PR	17-JUN-1999; 99US-0139453P.		
PR	17-JUN-1999; 99US-0139492P.		
PR	18-JUN-1999; 99US-0139454P.		
PR	18-JUN-1999; 99US-0139455P.		
PR	18-JUN-1999; 99US-0139456P.		
PR	18-JUN-1999; 99US-0139457P.		
PR	18-JUN-1999; 99US-0139458P.		
PR	18-JUN-1999; 99US-0139459P.		
PR	18-JUN-1999; 99US-0139460P.		
PR	18-JUN-1999; 99US-0139461P.		
PR	18-JUN-1999; 99US-0139462P.		
PR	18-JUN-1999; 99US-0139463P.		
PR	18-JUN-1999; 99US-0139750P.		
PR	18-JUN-1999; 99US-0139763P.		
PR	21-JUN-1999; 99US-0139817P.		
PR	22-JUN-1999; 99US-0139899P.		
PR	23-JUN-1999; 99US-0140353P.		
PR	23-JUN-1999; 99US-0140354P.		
PR	24-JUN-1999; 99US-0140695P.		
PR	28-JUN-1999; 99US-0140823P.		
PR	29-JUN-1999; 99US-0140991P.		
PR	30-JUN-1999; 99US-0141287P.		
PR	01-JUL-1999; 99US-0141842P.		
PR	01-JUL-1999; 99US-0142154P.		
PR	02-JUL-1999; 99US-0142055P.		
PR	06-JUL-1999; 99US-0142390P.		
PR	08-JUL-1999; 99US-0142803P.		
PR	09-JUL-1999; 99US-0142920P.		
PR	12-JUL-1999; 99US-0142977P.		
PR	13-JUL-1999; 99US-0143542P.		
PR	14-JUL-1999; 99US-0143624P.		
PR	15-JUL-1999; 99US-0144005P.		
PR	16-JUL-1999; 99US-0144085P.		
PR	16-JUL-1999; 99US-0144086P.		
PR	19-JUL-1999; 99US-0144325P.		
PR	19-JUL-1999; 99US-0144331P.		
PR	19-JUL-1999; 99US-0144332P.		
PR	19-JUL-1999; 99US-0144333P.		
PR	19-JUL-1999; 99US-0144334P.		
PR	19-JUL-1999; 99US-0144335P.		
PR	20-JUL-1999; 99US-0144352P.		
PR	20-JUL-1999; 99US-0144632P.		
PR	20-JUL-1999; 99US-0144884P.		
PR	21-JUL-1999; 99US-0144814P.		
PR	21-JUL-1999; 99US-0145086P.		
PR	21-JUL-1999; 99US-0145088P.		
PR	22-JUL-1999; 99US-0145089P.		
PR	22-JUL-1999; 99US-0145087P.		
PR	22-JUL-1999; 99US-0145192P.		
PR	23-JUL-1999; 99US-0145145P.		
PR	23-JUL-1999; 99US-0145218P.		
PR	23-JUL-1999; 99US-0145224P.		
PR	26-JUL-1999; 99US-0145276P.		
PR	27-JUL-1999; 99US-0145313P.		
PR	27-JUL-1999; 99US-0145318P.		
PR	28-JUL-1999; 99US-0145919P.		
PR	28-JUL-1999; 99US-0145951P.		
PR	02-AUG-1999; 99US-0146386P.		
PR	02-AUG-1999; 99US-0146388P.		
PR	02-AUG-1999; 99US-0146389P.		
PR	03-AUG-1999; 99US-0147038P.		
PR	04-AUG-1999; 99US-0147204P.		
PR	05-AUG-1999; 99US-0147302P.		
PR	05-AUG-1999; 99US-0147192P.		
PR	05-AUG-1999; 99US-0147360P.		
PR	06-AUG-1999; 99US-0147303P.		
PR	06-AUG-1999; 99US-0147416P.		

```

PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149428P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155113P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161952P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149428P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155113P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161952P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.
QY 1 MRLATKGQDIQECTTOIQVLKQVQPSVAQLRSTWDPAINLFLKMKGELEBQTKDKLEQ 60
DB 84 VQYAKREQEMAEKLSAVRDLKSQLKPSMCAARLLDPAIHEEFSRLKNLVEEKDKIKE 143
QY 61 AQNELSAWKFTPDSTGKKLMAKCRMLICENOEELGRLSQGRIAGLEAEALALOKKYSEEL 120
DB 144 LQDNIAAVTFTQSKNGKMLAKCETLOEENEIEHQAAEGKIHLAIAKLANQKSNL 203
QY 121 KSSQDELNDFIQLDEEVEGMSQSTILVLOQQKFTROQLAQYQQ 164
DB 204 RSQFGLYKHEELINDVRSNETVILQEKLEEKIEKIERVKX 247
RESULT 8
AAG41463
ID AAG41463 standard; protein; 330 AA.
XX
AC AAG41463;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 51591.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 04-MAY-1999; 99US-0132407P.
PR 05-MAY-1999; 99US-0132484P.
PR 06-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0132863P.
PR 14-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 18-MAY-1999; 99US-0134370P.
PR 19-MAY-1999; 99US-0134768P.
PR 20-MAY-1999; 99US-0134941P.
PR 21-MAY-1999; 99US-0135124P.
PR 24-MAY-1999; 99US-0135353P.
PR 25-MAY-1999; 99US-0135629P.
PR 27-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.

```

Query Match 17.5%; Score 282; DB 3; Length 278;
 Best Local Similarity 33.5%; Pred. No. 2, 9e-14;
 Matches 55; Conservative 51; Mismatches 58; Indels 0; Gaps 0;

PR	07-JUN-1999;	99US-0137724P.	PR	06-AUG-1999;	99US-0147416P.
PR	08-JUN-1999;	99US-0138094P.	PR	09-AUG-1999;	99US-0147493P.
PR	10-JUN-1999;	99US-0138540P.	PR	09-AUG-1999;	99US-0147935P.
PR	10-JUN-1999;	99US-0138847P.	PR	10-AUG-1999;	99US-0148171P.
PR	14-JUN-1999;	99US-0139119P.	PR	11-AUG-1999;	99US-0148319P.
PR	16-JUN-1999;	99US-0139452P.	PR	12-AUG-1999;	99US-0148341P.
PR	16-JUN-1999;	99US-0139453P.	PR	13-AUG-1999;	99US-0148565P.
PR	17-JUN-1999;	99US-0139492P.	PR	13-AUG-1999;	99US-0148684P.
PR	18-JUN-1999;	99US-0139454P.	PR	16-AUG-1999;	99US-0149368P.
PR	18-JUN-1999;	99US-0139455P.	PR	17-AUG-1999;	99US-0149175P.
PR	18-JUN-1999;	99US-0139456P.	PR	18-AUG-1999;	99US-0149426P.
PR	18-JUN-1999;	99US-0139457P.	PR	20-AUG-1999;	99US-0149722P.
PR	18-JUN-1999;	99US-0139458P.	PR	20-AUG-1999;	99US-0149723P.
PR	18-JUN-1999;	99US-0139459P.	PR	20-AUG-1999;	99US-0149929P.
PR	18-JUN-1999;	99US-0139460P.	PR	23-AUG-1999;	99US-0149902P.
PR	18-JUN-1999;	99US-0139461P.	PR	23-AUG-1999;	99US-0149930P.
PR	18-JUN-1999;	99US-0139462P.	PR	25-AUG-1999;	99US-0150566P.
PR	18-JUN-1999;	99US-0139463P.	PR	26-AUG-1999;	99US-0150884P.
PR	18-JUN-1999;	99US-0139750P.	PR	27-AUG-1999;	99US-0151065P.
PR	21-JUN-1999;	99US-0139817P.	PR	27-AUG-1999;	99US-0151066P.
PR	22-JUN-1999;	99US-0139899P.	PR	27-AUG-1999;	99US-0151068P.
PR	23-JUN-1999;	99US-0140333P.	PR	27-AUG-1999;	99US-0151080P.
PR	23-JUN-1999;	99US-0140334P.	PR	31-AUG-1999;	99US-0151303P.
PR	24-JUN-1999;	99US-0140695P.	PR	31-AUG-1999;	99US-0151438P.
PR	28-JUN-1999;	99US-0140823P.	PR	01-SEP-1999;	99US-0151930P.
PR	29-JUN-1999;	99US-0140991P.	PR	07-SEP-1999;	99US-0152363P.
PR	30-JUN-1999;	99US-0141287P.	PR	10-SEP-1999;	99US-0153070P.
PR	01-JUL-1999;	99US-0141842P.	PR	13-SEP-1999;	99US-0153758P.
PR	01-JUL-1999;	99US-0142184P.	PR	15-SEP-1999;	99US-0154018P.
PR	02-JUL-1999;	99US-0142055P.	PR	16-SEP-1999;	99US-0154039P.
PR	06-JUL-1999;	99US-0142330P.	PR	20-SEP-1999;	99US-0154779P.
PR	08-JUL-1999;	99US-0142803P.	PR	22-SEP-1999;	99US-0155139P.
PR	09-JUL-1999;	99US-0142920P.	PR	23-SEP-1999;	99US-0155486P.
PR	12-JUL-1999;	99US-0142977P.	PR	24-SEP-1999;	99US-0155655P.
PR	13-JUL-1999;	99US-0143542P.	PR	28-SEP-1999;	99US-0156458P.
PR	14-JUL-1999;	99US-0143624P.	PR	29-SEP-1999;	99US-0156596P.
PR	15-JUL-1999;	99US-0144005P.	PR	04-OCT-1999;	99US-0157117P.
PR	16-JUL-1999;	99US-0144085P.	PR	05-OCT-1999;	99US-0157753P.
PR	16-JUL-1999;	99US-0144086P.	PR	06-OCT-1999;	99US-0157865P.
PR	19-JUL-1999;	99US-0144332P.	PR	07-OCT-1999;	99US-0158029P.
PR	19-JUL-1999;	99US-0144331P.	PR	08-OCT-1999;	99US-0158232P.
PR	19-JUL-1999;	99US-0144332P.	PR	12-OCT-1999;	99US-0158369P.
PR	19-JUL-1999;	99US-0144333P.	PR	13-OCT-1999;	99US-0159293P.
PR	19-JUL-1999;	99US-0144334P.	PR	13-OCT-1999;	99US-0159294P.
PR	19-JUL-1999;	99US-0144335P.	PR	13-OCT-1999;	99US-0159295P.
PR	20-JUL-1999;	99US-0144332P.	PR	14-OCT-1999;	99US-0159329P.
PR	20-JUL-1999;	99US-0144332P.	PR	14-OCT-1999;	99US-0159330P.
PR	20-JUL-1999;	99US-0144884P.	PR	14-OCT-1999;	99US-0159331P.
PR	21-JUL-1999;	99US-0144814P.	PR	14-OCT-1999;	99US-0159637P.
PR	21-JUL-1999;	99US-0145086P.	PR	14-OCT-1999;	99US-0159638P.
PR	21-JUL-1999;	99US-0145088P.	PR	18-OCT-1999;	99US-0159584P.
PR	22-JUL-1999;	99US-0145087P.	PR	21-OCT-1999;	99US-0160741P.
PR	22-JUL-1999;	99US-0145089P.	PR	21-OCT-1999;	99US-0160767P.
PR	22-JUL-1999;	99US-0145192P.	PR	21-OCT-1999;	99US-0160768P.
PR	23-JUL-1999;	99US-0145145P.	PR	21-OCT-1999;	99US-0160770P.
PR	23-JUL-1999;	99US-0145218P.	PR	21-OCT-1999;	99US-0160814P.
PR	23-JUL-1999;	99US-0145224P.	PR	21-OCT-1999;	99US-0160815P.
PR	26-JUL-1999;	99US-0145276P.	PR	22-OCT-1999;	99US-0160980P.
PR	27-JUL-1999;	99US-0145913P.	PR	22-OCT-1999;	99US-0160981P.
PR	27-JUL-1999;	99US-0145918P.	PR	22-OCT-1999;	99US-0160989P.
PR	27-JUL-1999;	99US-0145919P.	PR	23-OCT-1999;	99US-0161404P.
PR	28-JUL-1999;	99US-0145917P.	PR	23-OCT-1999;	99US-0161405P.
PR	28-JUL-1999;	99US-0145955P.	PR	25-OCT-1999;	99US-0161406P.
PR	02-AUG-1999;	99US-0146386P.	PR	25-OCT-1999;	99US-0161359P.
PR	02-AUG-1999;	99US-0146388P.	PR	26-OCT-1999;	99US-0161360P.
PR	02-AUG-1999;	99US-0146389P.	PR	26-OCT-1999;	99US-0161361P.
PR	03-AUG-1999;	99US-0147038P.	PR	28-OCT-1999;	99US-0161920P.
PR	04-AUG-1999;	99US-0147204P.	PR	28-OCT-1999;	99US-0161922P.
PR	04-AUG-1999;	99US-0147202P.	PR	28-OCT-1999;	99US-0161933P.
PR	05-AUG-1999;	99US-0147192P.	PR	29-OCT-1999;	99US-0162142P.
PR	05-AUG-1999;	99US-0147260P.			
PR	06-AUG-1999;	99US-0147303P.			

Query Match 17.5%; Score 282; DB 3; Length 330;
Best Local Similarity 33.5%; Pred. No. 3.7e-14;
Matches 55; Conservative 51; Mismatches 58; Indels 0; Gaps 0;

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecularweight markers and as a food

CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 524 AA;

Query Match 10.3%; Score 166; DB 6; Length 524;
 Best Local Similarity 23.1%; Pred. No. 0.00016;
 Matches 84; Conservative 59; Mismatches 143; Indels 78; Gaps 15;

QY 8 QDIQECTIOQ-----YLKQVQPSVAQLRSTWDPAINLFFL-----KMKGELEQT 54
 DB 88 KDIALQERIEKREDTIQKQAREAQVNTSSNYIDAVLNADSLADAGRVOQMTWVKAN 147
 QY 55 KDKLEQAQNELSAWKFTPDSTQTKMLMAKRMILQENBELGRLQSQRIQAEL-ALQ 113
 DB 148 NDLMEQKQDKKA---VEDKKAEND--AKLKELAEQAALLESQ--KGDLLSKQADLNVLK 200
 QY 114 KKYSELKSSQDELNDFFIQLDEEVEGMSTILVLQQLKETRQQLAQ-----YQOQQ 166
 DB 201 TSLAEQNTAEKADKADLNQ--KAAEAQAEQIRQQLAQQAQQAQAEKAEQAQAE 259
 QY 167 SQASAPSTRRTTASEPVEQSEATSKDCSRLTNGPNSGSSSRQRTSGSGFHREGNTEDDF 226
 DB 260 AEAQATQASSTAQSSATESSATQSSMTTESSATQSSATESTT-----PESSTEESTA 314
 QY 227 PSSPFGNGKNSSEERTGSGGYVNLQSLAGYVESVDSPTGSENSLTHOSN-----DT 279
 DB 315 PES--SATEESTTAPESATEESTTVPESATEESTTVP-----ESSTEESTTAPTPTPT 369
 QY 280 DSSHD-----POEKAV-----SGKN-----RTVGRHRVQNG 307
 DB 370 DQSVDTGNGTSGSTTAPTPTPTPEQPKVTPAPAPSGSVNGAIAVAEAYKIGTPYVWGG 429
 QY 308 LDSS 311
 DB 430 KDPS 433

RESULT 13

AD95468
 ID AD95468 standard; protein; 525 AA.

XX
 AC AD95468;

DT 01-JAN-2004 (first entry)

XX DE B. faecium protein sequence SEQ ID 5095.

XX KW Vaccine; urinary tract infection; bacteraemia; endocarditis; wound;
 XX abdominal-pelvic infection.

XX OS Enterococcus faecium.

PN US583275-B1.

XX 24-JUN-2003.

XX 30-JUN-1998; 98US-00107532.

XX 02-JUL-1997; 97US-0051571P.

PR 14-MAY-1998; 98US-0085598P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Doucette-Stamm LA, Bush D;

XX WPI: 2003-799836/75.

DR N-PSDB; AOC91814.

XX New isolated nucleic acid derived from *Enterococcus faecium* encoding an
 PT Enterococcus faecium polypeptide useful for detection, prevention and
 PT treatment of a pathological condition resulting from a bacterial
 PT infection.

XX Example 1; SEQ ID NO 5095; 243pp; English.

XX The invention relates to an isolated nucleic acid derived from
 CC Enterococcus faecium encoding an Enterococcus faecium polypeptide having
 CC one of 10 fully defined sequences given in the (or comprising 40
 CC sequential nucleotides chosen from any of the nucleic acids, its
 CC complement or sequences hybridising to it). Also included are a
 CC recombinant vector comprising the nucleic acid operably linked to
 CC transcription regulatory element, a cell comprising the vector and a
 CC single-stranded probe comprising the nucleic acid. The nucleic acids are
 CC chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.
 CC The nucleic acids is useful for diagnosing pathological conditions
 CC resulting from *E. faecium* bacterial infection (e.g. urinary tract
 CC infection, bacteraemia, endocarditis, wounds and abdominal-pelvic
 CC infection) and for screening drugs such as agonists and antagonists. The
 CC nucleic acid is useful for recombinant production of *Candida albicans* -
 CC derived peptides or antisense polypeptides. Pharmaceutical compositions
 CC and vaccines containing the nucleic acid are useful for preventing or
 CC treating Enterococcus faecium infections. The present sequence represents
 CC one if the disclosed *E. faecium* proteins.

XX SQ Sequence 525 AA;

Query Match 10.3%; Score 166; DB 7; Length 525;

Best Local Similarity 23.1%; Pred. No. 0.00016;

Matches 84; Conservative 59; Mismatches 143; Indels 78; Gaps 15;

QY 8 QDIQECTIOQ-----YLKQVQPSVAQLRSTWDPAINLFFL-----KMKGELEQT 54
 DB 89 KDIALQERIEKREDTIQKQAREAQVNTSSNYIDAVLNADSLADAGRVOQMTWVKAN 148
 QY 55 KDKLEQAQNELSAWKFTPDSTQTKMLMAKRMILQENBELGRLQSQRIQAEL-ALQ 113
 DB 149 NDLMEQKQDKKA---VEDKKAEND--AKLKELAEQAALLESQ--KGDLLSKQADLNVLK 201
 QY 114 KKYSELKSSQDELNDFFIQLDEEVEGMSTILVLQQLKETRQQLAQ-----YQOQQ 166
 DB 202 TSLAEQNTAEKADKADLNQ--KAAEAQAEQIRQQLAQQAQQAQAEKAEQAQAE 260
 QY 167 SQASAPSTRRTTASEPVEQSEATSKDCSRLTNGPNSGSSSRQRTSGSGFHREGNTEDDF 226
 DB 261 AEAQATQASSTAQSSATESSATQSSMTTESSATQSSATESTT-----PESSTEESTA 315
 QY 227 PSSPFGNGKNSSEERTGSGGYVNLQSLAGYVESVDSPTGSENSLTHOSN-----DT 279
 DB 316 PES--SATEESTTAPESATEESTTVPESATEESTTVP-----ESSTEESTTAPTPTPT 370
 QY 280 DSSHD-----POEKAV-----SGKN-----RTVGRHRVQNG 307
 DB 371 DQSVDTGNGTSGSTTAPTPTPTPEQPKVTPAPAPSGSVNGAIAVAEAYKIGTPYVWGG 430
 QY 308 LDSS 311

Db 431 KDPS 434

RESULT 14
ABG93255
ID ABG93255 standard; protein; 511 AA.
XX
AC ABG93255;
XX
DT 21-NOV-2002 (first entry)
XX
DE C. albicans BAX-associated protein fragment SEQ ID 469.
XX
KW Bax; Bax-resistance; cytostatic; fungicide; immunosuppressive; virucide;
KW vasotropic; vaccine; gene therapy; proliferative disorder; cancer;
KW apoptosis; fungal; yeast; infection; autoimmune disease; ischaemia;
KW neurodegeneration; cell death.
XX
OS Candida albicans.
XX
PN WO200264766-A2.
XX
PD 22-AUG-2002.
XX
PF 21-DEC-2001; 2001WO-EP015398.
XX
PR 22-DEC-2000; 2000EP-00870318.
PR 04-JAN-2001; 2001EP-00870002.
PR 09-JAN-2001; 2001EP-00870003.
XX
PA (JANC) JANSSEN PHARM NV.
XX
PI Contreras RH, Eberhardt I, Luyten WHML, Reekmans RJ;
XX
DR WPI; 2002-667002/71.
DR N-PSDB; ABQ76521.
XX
PT New isolated nucleic acid representing a synthetic BAX-gene, useful as
PT medicament for treating, preventing and/or alleviating yeast or fungal
PT infections or proliferative disorders, or for preventing apoptosis in
PT certain diseases.
XX
PS Claim 36; Fig 2; 344pp; English.
XX
CC This invention describes a novel nucleic acid representing a synthetic
CC Bax gene. The Bax gene of the invention is useful for identifying Bax-
CC resistant yeast or fungi, identifying, or obtaining and identifying
CC Candida spp. sequences that are differentially expressed in a pathway
CC eventually leading to programmed cell death or identifying inhibitors or
CC inhibitor sequences of Bax-induced cell death. The products of the
CC invention have cytostatic, fungicide; immunosuppressive, virucide and
CC vasotropic activity and can be used in vaccines or for gene therapy. The
CC isolated nucleic acids, polypeptides, pharmaceutical compositions,
CC antisense molecules and antibodies are useful as medicaments or in
CC preparing a medicament for treating, preventing and/or alleviating
CC diseases associated with yeast or fungi or proliferative disorders, such
CC as cancer, or for preventing apoptosis in certain diseases. The compounds
CC or polypeptides, or the genetically modified organism are useful for
CC preparing a medicament for modifying the endogenic flora of humans and
CC other mammals. The vaccine is useful for immunising against yeast or
CC fungal infections. Apoptosis-related diseases include autoimmune disease,
CC ischaemia, diseases related with viral infections or neurodegenerations.
CC This sequence represents a polypeptide associated with the Bax gene
CC described in the disclosure of the invention
XX
SQ Sequence 511 AA;

Query Match 9.7%; Score 156.5; DB 5; Length 511;
Best Local Similarity 23.3%; Pred. No. 0.00087;
Matches 51; Conservative 48; Mismatches 132; Indels 87; Gaps 14;

44 FLNMKGELFTQDKLQQAQNELSAWKFPTDS-QTGKLMKAKRMILQENQELG----- 95

Db 4 FNOIRKLTVEQRRVOLLLEQTKTGNVTPEQIQQIDKEIECKAKPQQYQKGIYRNOL 63
QY 96 ---RQLSQGRIQAELALQKXSEELKSSQDE-----LNDPIIQLDVEEGMQST 144
Db 64 VLQAKAQQRQLQRRQQQQQQQQQQNNLKSQAQNNQNNQNNQNNQNNQNNQNNQNNQ 123
QY 145 -----ILVLOQLKETROQLAAYQOQQ-----SQASAFSTSR- 176
Db 124 GQTPQQQSFSPQLQAVQQQCFMNGSQQQQQQQQQQLNANKSALQQAQPSQAFPLVRP 183
QY 177 TTASEPVEQ-----SEATSKDCSRLTNGPSNGSSSRQRTS----- 211
Db 184 TPQSQPTAQAGVASQAATPGPRASQPTPSQTSRTGSALQQAQPSQASSTPQSQFQPLP 243
QY 212 GSGFHREGNTTDDFPSPGNGKNSNSSEERTGRGGGYVNLGAGYESVDSPGSGNS 271
Db 244 SESRHPSATTSKPLPQQPFGSGTAKS-PSVAATPAQNNGTVTARSA-----SPVAT--- 293
QY 272 LTHQSNDDTSSHDPOEE-KAVSGKGNRTVG-SRHVQNGL--DSSVNVQ 315
Db 294 -TTDSATTGSCPTPQQSRSRSGSSINLAGITRQSVPSLPISSSINVK 340

RESULT 15
AAR92751
ID AAR92751 standard; protein; 897 AA.
XX
AC AAR92751;
XX
DT 25-MAR-2003 (revised)
DT 09-JUN-1996 (first entry)
XX
DE Murine EGF receptor substrate, eps15.
XX
KW eps15; epidermal growth factor; EGF; receptor; substrate; TK;
KW tyrosine kinase; mitogenic signalling pathway; neoplasia; tumour;
KW diagnosis; therapy.
XX
OS Mus sp.
XX
PN US5487979-A.
XX
PD 30-JAN-1996.
XX
PF 22-JUL-1993; 93US-00095737.
XX
PR 25-AUG-1992; 92US-00935311.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Fazioli F, Difiore PP;
XX
DR WPI; 1996-105227/11.
DR N-PSDB; AAT16484.
XX
PT Poly-nucleotide(s) encoding eps15 - used to develop prods. for the
PT therapy, diagnosis and prognosis of neoplasia and related disorders.
XX
PS Claim 2; Col 35-40; 23pp; English.
XX
CC AAR92750 is the murine epidermal growth factor (EGF) receptor (EGFR)
CC substrate, eps15 (EGFR pathway substrate 15). The EGFR is not thought to
CC interact with known second messenger systems efficiently and for this
CC reason there is a need to ascertain the mechanism by which the EGFR
CC functions in mitogenesis. eps15 has been isolated and found to be
CC tyrosine phosphorylated by the EGFR tyrosine kinase and hence involved in
CC the regulation of mitogenic signals. eps15 polynucleotides can be used to
CC develop prods. for the therapy, diagnosis and prognosis of neoplasia and
CC other disorders connected with abnormal mitogenic signalling pathways.
CC eps15 also enhances cell response to mitogenic factors. (Updated on 25-
CC MAR-2003 to correct PF field.)
XX

Search completed: November 14, 2004, 12:33:14
Job time : 160 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2004, 12:33:24 ; Search time 40 Seconds
(without alignments)
767.329 Million cell updates/sec

Title: US-10-030-389-2

Perfect score: 1608

Sequence: 1 MRLATKGQDIQECTTQIYLV.....GSRHVQNGLDSSVNVQGSYL 319

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	282	17.5	330	2 T47576	FKBP12 interacting
2	156.5	9.7	507	2 S05542	hypothetical prote
3	156	9.7	897	2 A54696	EGF receptor subst
4	152.5	9.5	1589	2 T13606	hypothetical prote
5	151	9.4	2288	2 T29999	hypothetical prote
6	150	9.3	456	2 E86903	hypothetical prote
7	150	9.3	2251	2 T24490	hypothetical prote
8	148	9.2	543	1 QFMSL	neurofilament trip
9	145	9.0	1938	1 A40997	myosin heavy chain
10	144.5	9.0	859	2 D96502	hypothetical prote
11	144	9.0	461	2 JN0097	secreted 45k prote
12	144	9.0	1807	2 S03124	vitellogenin A2 pr
13	141.5	8.8	697	2 T03834	nuclear distributi
14	140.5	8.7	461	2 H84099	cell wall-binding
15	140	8.7	1031	1 A38713	kinesin heavy chai
16	139.5	8.7	869	2 A88710	protein C43G2.2 [i
17	139	8.6	1017	2 PC4035	cell-cycle-depende
18	138	8.6	355	2 C39725	hypothetical prote
19	138	8.6	577	1 A41289	moesin - human
20	138	8.6	2245	2 T18278	myosin heavy chain
21	137.5	8.6	782	2 A34218	Bic-D protein - fr
22	137	8.5	544	2 S07144	neurofilament trip
23	136	8.5	1837	2 T41023	probable nuclear p
24	135.5	8.4	472	2 S41720	intermediate filam
25	135.5	8.4	1509	1 A27224	myosin heavy chain
26	135.5	8.4	2139	2 A27224	myosin heavy chain
27	135	8.4	892	2 T50986	related to transcr
28	135	8.4	1436	2 S57238	forked protein 5.4
29	135	8.4	1449	2 S57237	forked protein 5.6

30 134.5 8.4 561 2 A31994 keratin 10, type I
31 134.5 8.4 782 2 T32155 hypothetical prote
32 134.5 8.4 1410 1 A57013 early endosome ant
33 134.5 8.4 1940 2 A59287 myosin heavy chain
34 134 8.3 645 2 A44881 keratin, 67k type
35 133 8.3 613 2 S48557 hypothetical prote
36 132.5 8.2 1054 2 D96519 myosin-like protei
37 132 8.2 790 2 T05576 vitellogenin 1 - m
38 132 8.2 930 2 T08588 vitellogenin 1 - m
39 132 8.2 1704 2 T43141 epidermal growth f
40 131.5 8.2 896 2 S43074 probable transcrip
41 131 8.1 662 2 T18233 myosin heavy chain
42 130.5 8.1 484 2 B33501 keratin, 54k type
43 130.5 8.1 526 1 K8BOVI hypothetical prote
44 130.5 8.1 967 2 S66852 smooth muscle myos
45 130.5 8.1 1938 2 JC5421

ALIGNMENTS

RESULT 1

T47576
FKBP12 interacting protein (PIP37) - Arabidopsis thaliana
N:Alternate names: Protein F24B22.130
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 09-Jul-2004
C:Accession: T47576
R:Blöcker, H.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.; Quettier, F.; Salanoubat, M.
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z23016
A:Accession: T47576
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-330 <BLO>
A:Cross-references: UNIPROT:Q9ZS28; EMBL:AL132957
A:Experimental source: cultivar Columbia; BAC clone F24B22
C:Genetics:
A:Map position: 3
A:Introns: 24/1; 60/2; 72/3; 102/1; 123/3; 149/3; 165/3; 183/3; 237/3; 261/1; 279/3
A>Note: F24B22.130

Query Match 17.5%; Score 282; DB 2; Length 330;

Best Local Similarity 33.5%; Pred. No. 5e-09;

Matches 55; Conservative 51; Mismatches 58; Indels 0; Gaps 0;

QY 1 MRLATKGQDIQECTTQIYLVQVQPSVAQLRSTWDPAINLFLKMKGELEQTKDKLEQ 60

Db 136 VOYAKREQEMAEKLSAVRDLKSLKPSNQARLLDPAIHFEFSRLKNLVEEKDKKIKE 195

QY 61 ACNELSAWKTPTDSOTGKKLMKCRMLIQENQELGRLSQGRIAQLAEALQKKYSBEL 120

Db 196 LQNTAAVTFPTQSKNGKMLMAKCTLOEENEIGHQAEKGKIHETAKLAWKQSQNAEL 255

QY 121 KSSQDELNDFFIQLDEEVGMQSTILVLOQQLKETRQQLAQYQQ 164

Db 256 RSQFEGLYKHMBELNDVRSNETVILQEKLEKEKSIERVKK 299

RESULT 2

S05542

hypothetical protein, 54K - Enterococcus faecium

C:Species: Enterococcus faecium

C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 09-Jul-2004

C:Accession: S05542

R:Fuerst, P.; Moesch, H.U.; Solioz, M.

Nucleic Acids Res. 17, 6724, 1989

A:Title: A protein of unusual composition from Enterococcus faecium.

A:Reference number: S05542; MUID:89385998; PMID:2780297

A:Accession: S05542

A:Molecule type: DNA

A:Residues: 1-507 <FUE>

A:Cross-references: UNIPROT:PI3692; GB:X16421; EMBL:M26048; NID:G43333; PIDN:CAA34442.1

[illegible]

```

3  LATKQDIOECTTQ---TOYLKQVOQPSVAQLRSTWDPAIN-LFPLKMKGELETKOKL 58
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
703 LATLKTEIGQVAGQNKVYGHLLTTVOQQOQATNLQGVNAAGKMWVMVSTTG---TPITL 758
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
59  EQAQNELSAWKFTPDSQTKGLMAKCMRLIOENQBLGQLSGRIAQLEAEALOKKYSE 118
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
759 QNGQTLHAATAGVDKXQQOQLQLFKQOILQCCQMLQCIIAIQMCCQQAQVCAQOQQQ 818
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
119 ELKSSQDELNDFIQLDEVEGMSQSTILVLOOLKETQQLAQYQ----- 163
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
819 QV-SQQQV---AQOQAVAQOQAVAQACQOQREQQOQVAAQAQHQALANATQOIL 874
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
164 -----QQOSQASAPSTSTTASEPVEQEATSKDCSR 195
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
875 QVAPNQFITSHQOQOQQLHNLIIQQLQQQAQVQVCAQOQOQOQOQOQOQOQOQ 934
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
196 LTNGPSNCSSRRQTSGSGFHRENTTDEDDFPSPGNGKS----SNSSEERTGCGGVV 252
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
935 IVQOQS-GATSQOQTSQOQOHHQSOQLQSSVFPVSSTTTPAGIATSSALQAALSAGAI 993
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
253 NQLS-AGYESVDSPTGSNSLTHQSN-----DTDSSHDPQSEKAVS 292
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
994 FOTAKPGCGSSSTPSSAVTINQSSITPLVSTTVAISICOACTOASQVHOHCILIS 1049
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

.....

```
RESULT 5
T29999
Hypothetical protein ZC8.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T29999
R:Ratcliffe, P.; Bradshaw, H.
Submitted to the EMBL Data Library, July 1996
A:Description: The sequence of C. elegans cosmid ZC8.
A:Reference number: Z20719
A:Accession: T29999
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2288 <LAT>
A:Cross-references: UNIPROT:Q23081; EMBL:U64862; PIDN:AA52624.1; GSPDB:GN00028; CESP:T29999
A:Experimental source: strain Bristol N2; clone ZC8
C:Genetics:
A:Gene: CESP:ZC8.4
A:Map position: X
A:Introns: 43/2; 80/3; 110/3; 153/2; 274/3; 738/2; 1249/1; 1392/3; 1454/1; 1602/1; 1686/3
Query Match 9.4%; Score 151; DB 2; Length 2288;
Best Local Similarity 21.6%; Pred. No. 0.99;
Matches 79; Conservative 54; Mismatches 137; Indels 96; Gaps 13;
QY 10 IQECTQIQVYLVKQV---QPSVAQLRSTWVDPAINLFFLMKGELEQTQKDKLEQAQNELS 66
DB 671 IEEKETQIRYSDDIRNICKDLDLREKY--DRVHTDNEKILGELSHAQAAHLAEQQKLK 728
QY 67 AKWFTDPS-----
DB 729 EIKIQRDDYQKQDEHARHFLDTRHKLETEIKGRQLEKNGKARNDELKLRQTTSIDYES 788
QY 83 KCRMLIQENQELGRQLS--QGRIQAQLEAEALQKQYSEEL-----KSSQDELNDFII 132
DB 789 QINLLRRHNDLDTTIKGHGQKITHLENELHRSRSGEIKLNDLNQRLQEKQDILNQ-KL 847
QY 133 OLDEEVEGMOSTILVLOOQLKFTQQLAQY-----QOQSQCAAPSTSRRTASE 181
DB 848 KLPQGDVQALKEITKLENELEKLNENKELVGEARDAANQOLSRANLLNKELEDTKQ 907
QY 182 PVEQSEATSKDC-----SRLTNGPSNGSSSRQRTSGSGFREGNTTDDFPSPFGNG 233
DB 908 DLKHSITVQKLEQDIDRLKLERLANIKGGRISRDSTGT-----DGAP-----G 953
QY 234 NKSSNSSEERTGRGGGYVNLQAGYVSDSPGSGENSLTHQNDTDSHDPQEEKAVSG 293
DB 954 DRSSVADPSRT-RGAAGSTVFVPA--EDIESRGGGEIDIP--SSGDVIHGRDGRDAGN 1010
QY 294 KGNRTV 299
DB 1011 RGHYTI 1016
RESULT 6
E86903
Hypothetical protein usp45 [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C:Species: Lactococcus lactis subsp. lactis
C>Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 09-Jul-2004
C:Accession: E86903
R:Boletín, A.; Winkler, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s8
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: E86903
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-456 <STO>
A:Cross-references: UNIPROT:Q9CDJ1; GB:AB005176; PID:g12725236; PIDN:AAK06327.1; GSPDB:G
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: usp45
```

```
Query Match 9.3%; Score 150; DB 2; Length 456;
Best Local Similarity 24.9%; Pred. No. 0.18;
Matches 73; Conservative 47; Mismatches 103; Indels 70; Gaps 13;
QY 2 RLATKGODIOBCTTIQYLVKQVQPSVAQLRSTWVDPAINLFFL-----KMKGE 50
DB 88 QIATLINESIARTKTLB--QAQSAQVNSATYMDAVNSKSLTDVIQKTAIATVSSA 145
QY 51 LEQTKDKLEQAQNELSAWKFTPDQTKK---LMAKCRMLIOENQELGQLSGRIACL 106
DB 146 NKQMLEQKEKEKELS-----QKSETVKKNYQVSLSQSLDSQAQELTSCQAEKVATL 200
QY 107 BAELALQKQYSEELKSSQDELNDFIIOLDEEVEGMOSTILVLOOQLKFTQQLAQYQOQ 166
DB 201 -----NYQATITATAQDKKQSL---LDEKA-----AAEKAQAAKQJAYEAQ 241
QY 167 SQAS---APSTSRIT-----ASEEVEQSEATSKDCSRLTNGPSNGSSSRQRTSGFPHRE 218
DB 242 KEAKAQAAATAATTKAEEATSVSSQA--SQSSSSSSNTSSNTSSSSSSSN--- 296
QY 219 GNTTDDFPSPFGNGKSSNSSEERTGRGGGYVNLQAGYVSDSPGSGENS 271
DB 297 -----SSSSSSSSSSSS---NGGGSTNTGNNAAG-----TGNTGGSSS 333
RESULT 7
T24490
Hypothetical protein T05A10.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T24490
R:Sulston, J.
Submitted to the EMBL Data Library, November 1995
A:Reference number: Z19898
A:Accession: T24490
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2251 <WTL>
A:Cross-references: UNIPROT:Q22190; EMBL:Z68108; PIDN:CAA92133.1; GSPDB:GN00028; CESP:T.
A:Experimental source: clone T05A10
C:Genetics:
A:Gene: CESP:T05A10.1
A:Map position: X
A:Introns: 188/3; 240/3; 420/1; 570/3; 596/1; 732/3; 778/3; 851/3; 1359/2; 1394/2; 1434
Query Match 9.3%; Score 150; DB 2; Length 2251;
Best Local Similarity 20.5%; Pred. No. 1.1;
Matches 82; Conservative 68; Mismatches 144; Indels 106; Gaps 15;
QY 4 ATKGDQIQ-----ECTTQIQYLVKQVQPSVAQLRSTWVDPAINL-----FFLKM 47
DB 248 AQCAQHVSQRQVQPSQSQVQALQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ 307
QY 48 KGELEQTQKDKLEQAQNELSAWKFTPD-----SOTGKLMKCRMLIOEN 91
DB 308 QOQLHCQRAAQAQQAQNAQNSQQRPSVASTPALSTPOLNDLTQTMQLOQO--QLLQOQ 366
QY 92 QELGQLSGRIQAQLEAEALQKQYSEELKSSQDELNDFI--IQLEDEVEG---MQSTIL 146
DB 367 QAQAQQAQQAQOQAL-AQAQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ 425
QY 147 VLOQQLKTRQQLAQYQOQOQO----- 168
DB 426 RLQAQMLNEQQRMQLQNAQRAHQHRLILSSTPAPRGGITMTGPTIGIARREQPVSTVAV 485
QY 169 ASAPSTSR--TASEEVEQSEATSKDCSRLTNGPSNGSSSRQRTSGSGFPHRENTTDDFP 227
DB 486 TTAQAAVTPVAVPMKQNSNPMN-----PSSTSTASATSSHQILAPLSKPLEQ 538
QY 228 SS---PFGNGKSSNS-----SEERTGRGGGYVNLQAGYVSDSPGSGENSLTHQSN- 277
DB 539 SSSKAASSSGNESMSDHISRIISENEVILQGDPIVRKRPYHRQI----GAQSSVDHDSNS 594
```

QY 278 --DTSDHPDQEKAVSGKN-----RTVGRHVQNGLDS 310
 Db 595 GGSRTSPGPKDSRMLQAASRSQSLFELSGSKHFGSLTS 634

RESULT 8

QPMSL

neurofilament triplet L protein - mouse

N:Alternate names: 69K neurofilament protein; NF-L(low) protein; type IV IF protein
 C:Species: Mus musculus (house mouse)

C>Date: 31-Mar-1988 #sequence revision 30-Sep-1993 #text_change 09-Jul-2004

C:Accession: A25227; A26562; A43772; A41012; I55316

R:Lewis, S.A.; Cowan, N.J.

Mol. Cell. Biol. 6, 1529-1534, 1986

A:Title: Anomalous placement of introns in a member of the intermediate filament multigene

A:Reference number: A25227; MUID:87064433; PMID:3785173

A:Accession: A25227

A:Molecule type: DNA

A:Residues: 1-543 <LEW>

A:Cross-references: UNIPROT:P08551; GB:M13016; NID:G200023; PIDN:AAA39810.1; PID:G387492

A:Note: the authors translated the codon GGC for residue 5 as Ala, ACA for residue 88 as

1 as Glu

R:Lewis, S.A.; Cowan, N.J.

J. Cell Biol. 100, 843-850, 1985

A:Title: Genetics, evolution, and expression of the 68,000-mol-wt neurofilament protein:

A:Reference number: A26562; MUID:8511334; PMID:3919033

A:Accession: A26562

A:Molecule type: mRNA

A:Residues: 242-543 <LE2>

A:Cross-references: GB:X02165

A:Experimental source: brain

R:Julien, J.P.; Meyer, D.; Flavell, D.; Hurst, J.; Grosveld, F.

Brain Res. Mol. Brain Res. 1, 243-250, 1986

A:Title: Cloning and developmental expression of the murine neurofilament gene family.

A:Reference number: A43772

A:Accession: A43772

A:Molecule type: mRNA

A:Residues: 1-5, 'Y', '7-8, 'Y', '10-64, 'M', '66-72, 'L', '74-98, 'D', '100-194, 'R', '196-202, 204-239, 'Y

A:Cross-references: GB:M20480; NID:G200037; PIDN:AAA39814.1; PID:G200038

A:Note: the authors translated the codon CGC for residue 195 as Ala

R:Shag, R.K.; Nixon, R.A.

J. Biol. Chem. 266, 18861-18867, 1991

A:Title: Identification of Ser-55 as a major protein kinase A phosphorylation site on the

A:Reference number: A41012; MUID:92011653; PMID:1717455

A:Accession: A41012

A:Molecule type: protein

A:Residues: 52-57 <STH>

R:Nakahira, K.; Ikenaka, K.; Wada, K.; Tamura, T.

J. Biol. Chem. 265, 19786-19791, 1990

A:Title: Structure of the 68-kDa neurofilament gene and regulation of its expression.

A:Reference number: I55316; MUID:91060592; PMID:2246261

A:Accession: I55316

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-5, 'Y', '7-8, 'Y', '10-28 <RES>

A:Cross-references: GB:M55423; NID:G200027; PIDN:AAA39812.1; PID:G554245

C:Comment: This is the most abundant of the three neurofilament proteins and, as the oth

C:Genetics:

A:Introns: 349/3; 391/2; 498/1

C:Superfamily: cytoskeletal keratin

C:Keywords: coiled coil; intermediate filament

F:2-72/Domain: head <HED>

F:94-125/Domain: coil 1a, alpha-helical rod #status predicted <R1a>

F:126-138/Region: linker 1

F:139-234/Domain: coil 1b, alpha-helical rod #status predicted <R1b>

F:235-256/Region: linker 12

F:257-272/Domain: coil 2a, alpha-helical rod #status predicted <R2a>

F:273-281/Region: linker 2

F:282-401/Domain: coil 2b, alpha-helical rod #status predicted <R2b>

F:404-543/Domain: tail <TAI>

F:404-444/Region: tail subdomain a

F:445-543/Region: tail subdomain b

Query Match 9.2%; Score 148; DB 1; Length 543;
 Best Local Similarity 19.3%; Pred. No. 0.28;

Matches 69; Conservative 55; Mismatches 164; Indels 70; Gaps 9;

QY 4 ATKGDQ-----IQECTTQIQYLKQVQPSVAQLR-----STMVD-----PA 39

Db 195 ARKGDEAALAAELAELEKIDSLMDEIAFLKKVHEEIAELQALQIAQISYEMDVSKPD 254

QY 40 INLFELKMKVGELETKDKLEQAQNELSAWKTTPDSQTGKLMKRCMLIQNQLGRLS 99

Db 255 LSAAALKDIPRAQYELAAKQMNQAEWFKSRFTVLTESAKNTDAVRAKDEVSE-SRLL 313

QY 100 QGRTAQLAEALQKYSEELKSSQDELNDFIOLDDEVEGQSTILVLOQLKETRQQL 159

Db 314 KAKTLEIEACRGMNEALEKQLQLELEDQN-----ADISAMQDTINKLENELASTKSEM 366

QY 160 AQY---QQQSCASAPSTRTTASEPVEQSEATS---KDCSRLTNGSPNGSSSRQRTSGS 213

Db 367 ARYLKEYQDILNVKALDIEIAAYRKLEGETRLSTVSGSIISGYSQSSQVFGSGAYS 426

QY 214 GFHR-----EGNTTEDDFPSPGNGKNSNSEERT 244

Db 427 GLQSSSYLMSARSPPAYVTSHVQEQTEVEETIATKAEEAKDEPPSGEAEEREKEKEE 486

QY 245 GRGGSGYVQLSAGYVESVDSPTGSENLSLTHQSNLTDSSHDPQBEKAYSGKNTVGSF 302

Db 487 GEEEGAEDEEAAKDESEDTKEEEEG---EGEEEDTKSEEEKEEAGEEQVAKK 541

RESULT 9

A40997

Myosin heavy chain, striated adductor muscle - scallop (Aequipecten irradians)

N:Contains: myosin A1pase (EC 3.6.4.1)

C:Species: Aequipecten irradians

C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004

C:Accession: A40997; S13557

R:Nyitra, L.; Goodwin, E.B.; Szent-Gyorgyi, A.G.

J. Biol. Chem. 266, 18469-18476, 1991

A:Title: Complete primary structure of a scallop striated muscle myosin heavy chain. Seq

A:Reference number: A40997; MUID:92011595; PMID:1917970

A:Accession: A40997

A:Molecule type: mRNA

A:Residues: 1-1938 <NYI>

A:Cross-references: UNIPROT:P24733; GB:X55714; NID:G5611; PIDN:CAA39247.1; PID:G5612

C:Superfamily: myosin heavy chain; myosin motor domain homology

C:Keywords: actin binding; ATP; coiled coil; hydrolase; muscle contraction; nucleotide b

F:86-763/Domain: myosin motor domain homology <MMOT>

F:176-183/Region: nucleotide-binding motif A (P-loop)

F:547-586/Region: actin binding #status predicted

F:653-675/Region: actin binding #status predicted

F:836-1938/Domain: coiled coil #status predicted <COI>

F:836-1276/Region: S2

F:1277-1938/Region: light meromyosin

F:182/Binding site: ATP (Lys) #status predicted

F:693/703/Active site: Cys #status predicted

Query Match

Best Local Similarity 9.0%; Score 145; DB 1; Length 1938;

Matches 72; Conservative 63; Mismatches 146; Indels 60; Gaps 11;

QY 3 LATHGQD-----IQECTTQIQYLKQVQPSVAQLRSTWDPAINLFF-LKKKGELEQTKDK 57

Db 1189 LRKHGDAANEMADQVDQLQKVKSKLEKDKKDLKREMDLESQMTNHNKNGSEVMKQ 1248

QY 58 LEQAQNELSAWKTTPDSQTGKLMKRCMLIQNQLGRLQ---SQGRTAQLAEALQKK 115

Db 1249 FESQMSDLNA-RLDSQSRINELQSKSRLOANSDLTLQLEDAEHRVSVLSKEKQLSS 1307

QY 116 YSELKSSQDELNDFIOLDDEVEGQSTILVLOQLKETRQQLAQYQQQSQAS-----170

Db 1308 QLEDAARRSLEETRAKSKLQNEVRNHDMDATREQLQEBEQESKSDVQRLSKANNEIQ 1367

QY 171 -----APSTSR-----TASEPVEQSEATSKDCSRLTNGPS-----201


```

Db      1368  WRSKFESEGANTELEUDDQKRLKLSFAGDTTEAANAKCSALEKAKSRLOQELEDMSI 1427
QY      202    ----NGSSSRORTSGSPHREGNTTDDFPSPGNGKNKSSNBERTGSGGYVNOQS 256
Db      1428  EVDRAANASVOMEKKQAFDK--TTAE--WQAKVNSLQSELENSQKSGRYSAELYRIK 1482
QY      257  AGVESVDSPGTS---ENSJLTHQSNDDTSSHDPQEEKAVSGK 294
Db      1483  ASIEEYQDSIGALRRN-----KNLADEIHDLTDQLSEGR 1518

RESULT 10
D96502
Hypothetical protein F28H19.10 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: D96502
R:Theologian, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizlar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: D96502
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1859 <STO>
A:Cross-references: GB:AE0051173; NID:g7523675; PIDN:AAF63115.1; GSPDB:GN00141
C:Genetics:
A:Gene: F28H19.10
A:Map position: 1

```

```

secreted 45K protein precursor - Lactococcus lactis
C;Species: Lactococcus lactis
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999
C;Accession: JN0097
R;van Asselodonk, M.; Rutten, G.; Oteman, M.; Siezen, R.J.; de Vos, W.M.; Simons, G.
Gene 95, 155-160, 1990
A;Title: Cloning of usp45, a gene encoding a secreted protein from Lactococcus lactis s
A;Reference number: JN0097; UID:91071599; PMID:2123812
A;Accession: JN0097
A>Status: preliminary
A;Molecule type: DNA
A;Residues: 1-461 <VAN>
A;Cross-references: GB:M35374

Query Match          9.0%; Score 144; DB 2; Length 461;
Best Local Similarity 22.0%; Pred.No. 0.39;
Matches 65; Conservative 53; Mismatches 124; Indels 54; Gaps 9;

QY      2 RLATKGQDIQECTIQYLKQVQPVSQAQLRSWTWPPAINLFFL-----RKXGE 50
         ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
DB    88 QIATLNESIKERTKLE--AARSAQVNSSATNMVDNVNKSGLTDVIQKVTAIAIVSSA 145

QY      51 LEQTGDKLEAQNELSAWKETPPDSOTGKK---LMAKCRMLIQENQELGRQLSGRIAQL 106
         ::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
DB   146 NKQMLESQEKEQEKELS-----QKSETVKKNYNFVLSQSLSQAQELTSQQAECLKVATL 200

QY     107 EAELALQKYSEBELKSSQDELNFIIQLDEEVGMOSTILVLQQLKKTETRCQLAQYQQQQ 166
         ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
DB   201 -----NYQAIATAQDKKQAL---LDEKA-----AAEKAQAQAAKQAAYEAQQ 241

QY     167 SQASAPSTRTTASEPVEQGEARTSKDCSRUTNPFGNSSSRRQRTSGSGFHREGNITTEDDF 226
         ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
DB   242 KEAQAQAQAASTA-----AFAKAVEAATSSASASSQAQPQVSTS---TDMTTSNAS 288

QY     227 PSAPGNGKNGKSSSEBRTGTGGGYNNOLSGAVESVDPTGSSENSLTHQSNDDTSS 282
         ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
DB   289 ASNSSNSSNSSSSSSSSSSSSSSSSSSSSSSSAGGVNTSGTGTGTTGGSGINSS 344


RESULT 12
S03124
Vitellogenin A2 precursor - African clawed frog
N;Contents: lipovitellin I; lipovitellin II; phosvette I; phosvette II; phosvitin
C;Species: Xenopus laevis (African clawed frog)
C;Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 09-Jul-2004
C;Accession: S03124; I51570; B23876; PC2160; SI0625
R;Gerber-Huber, S.; Nardelli, D.; Haefliger, J.A.; Cooper, D.N.; Givel, F.; Germond, J.
Nucleic Acids Res. 15, 4737-4760, 1987
A;Title: Precursor-product relationship between vitellogenin and the yolk proteins as d
A;Reference number: S03124; UID:87259958; PMID:3601655
A;Accession: S03124
A;Molecule type: DNA
A;Cross-references: UNIPROT:P18709; EMBL:Y00354; NID:g65210; PIDN:CAA68433.1; PID:g49388
A;Note: 662-Ile, 958-Thr, and 1572-Lys were also found
R;Nardelli, D.; van het Schip, F.D.; Gerber-Huber, S.; Haefliger, J.
J. Biol. Chem. 262, 15377-15385, 1987
A;Title: Comparison of the organization and fine structure of a chicken and a Xenopus l
A;Reference number: I50441; UID:89058863; PMID:3680202
A;Accession: I51570
A>Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1807 <VAR>
A;Cross-references: GB:M18061; NID:g214877; PIDN:AAA49982.1; PID:g214878
R;Walker, P.; Brown-Juedi, M.; Germond, J.E.; Wanli, W.; Meijlink, F.C.P.W.; van het SC
EMBO J. 2, 2271-2279, 1993
A;Title: Sequence homologies within the 5' end region of the estrogen-controlled d vica
A;Reference number: A23876; UID:84131940; PMID:6199194
A;Accession: B23876
A>Status: preliminary
A;Molecule type: DNA
A;Residues: 1-35, 37-71 <WAL>
R;Montorzi, M.; Falchuk, K.H.; Vallee, B.L.
```

Biochem. Biophys. Res. Commun. 200, 1407-1413, 1994
 A>Title: Xenopus laevis vitellogenin is a zinc protein.
 A:Reference number: PC2160; MUID:94242001; PMID:8185593
 A:Accession: PC2160
 A:Molecule type: protein
 A:Residues: 16-25 <MON>
 A>Note: lipovitelin I
 R:Wallace, R.A.; Hoch, K.L.; Carnevali, O.
 J. Mol. Biol. 213, 407-409, 1990
 A>Title: Placement of small lipovitelin subunits within the vitellogenin precursor in x
 A:Reference number: S10624; MUID:90278951; PMID:2352275
 A:Accession: S10625
 A:Molecule type: protein
 A:Residues: 1291-1302 <WA2>
 A>Note: lipovitelin II
 C:Comment: This protein contains a stoichiometric amount of zinc.
 C:Superfamily: vitellogenin
 C:Keywords: egg yolk; glycoprotein; oocyte; phosphoprotein; zinc
 F:1-15/Domain: signal sequence #status predicted <SIG>
 F:16-1807/Product: vitellogenin A2 #status experimental <MAT>
 Query Match 9.0%; Score 144; DB 2; Length 1807;
 Best Local Similarity 21.5%; Pred. No. 1.9;
 Matches 64; Conservative 58; Mismatches 119; Indels 56; Gaps 9;
 QY 22 QVQPSVAQLRSTMVDPAINLFLFKMGLEQTKDKLEQAQNELSAWKFTPDSCGKLM 81
 DB 1004 QCKTHNARSNTIFYQAVGEHDFKLTMPKPAHGALEKQLQITA-----GPKAA 1054
 QY 82 AKRMLIQENQELGRQLSQCRQLAQLAELALQKYSBELKSSODELNDFIQLDEEVEGM 141
 DB 1055 SKIMGLVEVEGTEGPM-----ETAVTKR-----LK-----MILGIDESKDT 1093
 QY 142 QSTLVQLQQLKETR-----QQLAQYQQQSQASAPTSRTTASEPVEQSATSQDCS 194
 DB 1094 NETALYSKQKKKKNINRLDAEVEARKQQLSSSSSSSSSSSSSSSSSSSSSSSSSS 1148
 QY 195 RLTPGPNSSS-----SRQTSGSGFREGNTTDDFPSSPGNGKNSNSSEERTGRGG 248
 DB 1149 SSSSSPSSSSSSSVYSKSKREHNPHQRESS-----SSSQEQNKMLQENR---KHG 1200
 QY 249 SGYNQLSAGYVDSFTGNSLTHOSNLTDSHDPQEKAVSGKGNRTVGRHVQ 305
 DB 1201 QKGMSS 1257
 RESULT 13
 T03834
 nuclear distribution protein Roll - Neurospora crassa
 C:Species: Neurospora crassa
 C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004
 R:Minke, P.F.; Tinsley, J.H.; Bruno, K.S.; Plamann, M.
 C:Accession: T03834
 submitted to the EMBL Data Library, July 1997
 A:Description: Neurospora ro-10 and ro-11 genes encode novel proteins required for nucle
 A:Reference number: Z15107
 A:Accession: T03834
 A>Status: preliminary; translated from GS/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-697 <MIN>
 A:Cross-references: UNIPROT:O13335; EMBL:AF015560; NID:G23553162; PIDN:AAB69381.1; PID:G2
 C:Genetics:
 A:Gene: ro-11
 A:Introns: 77/3
 Query Match 8.8%; Score 141.5; DB 2; Length 697;
 Best Local Similarity 22.8%; Pred. No. 0.87;
 Matches 76; Conservative 59; Mismatches 128; Indels 71; Gaps 13;

QY 18 QYLKQVQOPS---VAQLRSTM---VDPAINLFLFKMG---ELEOTKDKLEQAQNELSA-- 67
 DB 32 QELKEFQSSKELEAELEKDDAADKRERALQOKAEGLSYVEWKRKYKESKEANAAQ 91
 QY 68 ----WKFTPDSCGKLMKACRMLIQENQELGRQLSQCRQLAQLAELALQKYS----- 117
 DB 92 SALEKEITALRETNLTQLKLRDIEVANDPERQ-ARNTSSLE---DLESKYNVAIERA 147
 QY 118 ----BELKSSODELNDFIIQ---LDEEVEGQSTILVLOQOLKETRQQLAQYQQQSOAS 170
 DB 148 VMMEEEIKIGOEERELRVEAQRUEELSDIKIEAILQSKLRKHQARGH/TQITTTIAP 207
 QY 171 APSTSRITASEFV-----EQSEATSKDCSRLTNGSPNSSSSRORTSGSGFREGNTTDD 224
 DB 208 APASPLSTASSPLVSTPPDTKSLSDTLSEVQDPPSPMS--DASLGKGLRASRSTPVK 265
 QY 225 DFPSSPG-----NGKSNSSSEERTGRGGSGYNNQLSAGYVDSFTGNSLTHQ 275
 DB 266 QTASRPGGCRTPKTSISKAAAKSAQATHKANOSF-----SENNITPK 309
 QY 276 SNTDSSHDPQEKAVSGKGN-----RTVGSR 302
 DB 310 PKPLSSSTSSQSNRPNRPNFNTNSYPMVPTPSSR 343
 RESULT 14
 H84099
 cell wall-binding protein BH3600 [imported] - Bacillus halodurans (strain C-125)
 C:Species: Bacillus halodurans
 C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 09-Jul-2004
 C:Accession: H84099
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
 Nucleic Acids Res. 28, 4317-4331, 2000
 A>Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
 A:Reference number: A33650; MUID:20512582; PMID:11058132
 A:Accession: H84099
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-461 <STO>
 A:Cross-references: UNIPROT:Q9K6X4; GB:AF001519; GB:BA000004; NID:G10176109; PIDN:BA0073
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: BH3600
 Query Match 8.7%; Score 140.5; DB 2; Length 461;
 Best Local Similarity 20.4%; Pred. No. 0.61;
 Matches 68; Conservative 65; Mismatches 113; Indels 87; Gaps 15;
 QY 2 RLATKGODIQCTTQIQVLYK---OVQOPSVAQLRSTMVDPAINLF-----FLKMGKEL 51
 DB 86 KIQEKREIEVQAEIEBELKEQIEILBERTAERDELKDRARANYQNGSIDYLEVLGA 145
 QY 52 ECTKDKLEQAQNELSAWKFTPDSCG-----KLMKACRMLIQENQELGRQLSQCR 103
 DB 146 KSGDFLDR-----VSALSVAEQDQGLEAHIEDHRLLEAKAQVEKLE-----TLEGHL 197
 QY 104 AQLE---AELALQKYS---ELKSSODELNDFIQLDEEVEGQSTILVLOQOLKETR 156
 DB 198 VELNLMAQLQEQKEKQKMGELASREDEHLGDLSELENDEE-----LLRQEQKALQ 250
 QY 157 QQLAQYQQQSQASAPTSRTTASEPVEQSATSQDCSRLTNGSPNSSSSRORTSGSGGFH 216
 DB 251 EVELWKQKEERKA--AEKAAAEAAQAQAS-----SGGGSGNSDSDGS--- 295
 QY 217 REGNTTDDFPSSPGNGKNSNSSEERTGRGGSGYNNQLSAGYVDSFTGNSLTHOS 276
 DB 296 NSGTTSSNGSGSGGGGGETGSPVSS-----SGSGFMRPATG---DISSPFGYR---THP 344
 QY 277 NDTDSSHDPQEKAVSGKGNRTVGRHVQNGLD 309
 DB 345 -----VTGQRKLHAGID 356

Search completed: November 14, 2004, 12:46:19
Job time : 43 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2004, 12:36:39 ; Search time 141 Seconds
(without alignments)
800.484 Million cell updates/sec

Title: US-10-030-389-2

Perfect score: 1608

Sequence: 1 MRLATKGDIQECTTQIQYL.....GSRHVQGLDSVNVQGSVL 319

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1568699 seqs, 353819137 residues

Total number of hits satisfying chosen parameters: 1568699

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

```
1: /cgn2_6/ptodata1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata1/pubpaa/PT_NEW_PUB.pep.*
3: /cgn2_6/ptodata1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata1/pubpaa/PTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata1/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata1/pubpaa/US09_PUBCOMB.pep.*
12: /cgn2_6/ptodata1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata1/pubpaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata1/pubpaa/US10_PUBCOMB.pep.*
15: /cgn2_6/ptodata1/pubpaa/US10_PUBCOMB.pep.*
16: /cgn2_6/ptodata1/pubpaa/US10_PUBCOMB.pep.*
17: /cgn2_6/ptodata1/pubpaa/US10_NEW_PUB.pep.*
18: /cgn2_6/ptodata1/pubpaa/US11_NEW_PUB.pep.*
19: /cgn2_6/ptodata1/pubpaa/US60_NEW_PUB.pep.*
20: /cgn2_6/ptodata1/pubpaa/US60_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	357	22.2	151	9	US-09-840-787-1
2	351	21.8	188	17	US-10-425-115-191548
3	272	16.9	209	16	US-10-767-701-39876
4	269.5	16.8	376	17	US-10-425-115-307181
5	255.5	16.5	265	15	US-10-425-114-57475
6	264	16.4	343	15	US-10-424-599-240261
7	262.5	16.3	195	17	US-10-425-115-203631
8	256	15.9	213	15	US-10-425-114-48881
9	256	15.9	336	16	US-10-437-963-196776
10	256	15.9	352	14	US-10-310-154-593
11	247	15.4	190	15	US-10-425-114-55340
12	238	14.8	299	17	US-10-739-930-9291
13	238	14.8	310	15	US-10-424-599-240263

```
14 234 14.6 300 15 US-10-424-599-240259
15 218.5 13.6 165 15 US-10-425-114-42229
16 203 12.6 146 15 US-10-424-599-208116
17 166 10.3 524 15 US-10-282-122A-57658
18 156.5 9.7 511 16 US-10-451-467A-468
19 152.5 9.5 1031 9 US-09-815-242-10932
20 151 9.4 2286 14 US-10-369-493-6774
21 150 9.3 431 16 US-10-383-930-30
22 149 9.3 431 16 US-10-383-930-33
23 146 9.1 431 16 US-10-383-930-29
24 144 9.0 716 14 US-10-114-774-20
25 142.5 8.9 689 13 US-10-108-605-305
26 142.5 8.9 1061 15 US-10-282-122A-57375
27 142.5 8.9 3161 16 US-10-437-963-176140
28 141.5 8.8 432 16 US-10-383-930-31
29 141 8.8 294 15 US-10-264-049-2795
30 141 8.8 1742 15 US-10-615-383-4
31 141 8.8 1742 16 US-10-690-184-4
32 141 8.8 1742 16 US-10-689-082-4
33 140 8.7 1031 14 US-10-080-608A-24
34 140 8.7 1031 14 US-10-370-685-113
35 139 8.6 1017 16 US-10-408-765A-158
36 138.5 8.6 542 15 US-10-205-331-57
37 138 8.6 577 14 US-10-236-031B-54
38 138 8.6 577 16 US-10-360-849A-27
39 138 8.6 577 16 US-10-408-765A-453
40 138 8.6 577 16 US-10-408-765A-454
41 138 8.6 813 9 US-09-964-899-25
42 138 8.6 1165 14 US-10-021-660-126
43 138 8.6 1165 15 US-10-211-462-89
44 138 8.6 868 16 US-10-437-963-123248
45 137 8.5
```

ALIGNMENTS

RESULT 1

US-09-840-787-1

; Sequence 1, Application US/09840787

; Patent No. US2002058264A1

; GENERAL INFORMATION:

; APPLICANT: Lal, Preeti

; Hillman, Jennifer L.

; Bandman, Olga

; Shah, Purvi

; Au-Young, Janice

; Yue, Henry

; Guegler, Karl J.

; Corley, Neil C.

; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES

; NUMBER OF SEQUENCES: 98

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/840,787

; FILING DATE: 23-Apr-2001

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 09/518,865

; FILING DATE: <Unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J.

; NAME: Billings, Lucy J.

US-10-767-701-39876

DB 233 NVAANFTSSKLGKMLAKCRTLOEENEIGAMASEGKIHELGMKIAVLKSNQNELRNQ 292
QY 124 QDELNDFFIQLDEEVEGMSQSTILVLQQLKQKTRQQLAQVQOOQ-SQASAPSTSRRTTASEP 182
DB 293 FDVLYKHMGGVNDVRSNELVAILQEELEAKELEVLRLKEETLSQKGTPEAPAPAP 352
QY 183 VEQSEATSKD 192
DB 353 VEEGDGAGND 362

RESULT 5

US-10-425-114-57475
; Sequence 57475, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 57475
; LENGTH: 285
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZWFLB73213A07_FLI.pep
US-10-425-114-57475

Query Match 16.5%; Score 265.5; DB 15; Length 265;
Best Local Similarity 31.7%; Pred. No. 3.5e-10;
Matches 65; Conservative 47; Mismatches 80; Indels 13; Gaps 3;

QY 4 ATKQDIOECTTQIYLYKQVQPSVAQLRSTWVDPAINLFLKMGKELETKDKLEQAQN 63
DB 69 AKRQEIAELKSAVRDLKQVLPSPMQTRRLDPAIHFEFTRLKLVKEKIKELQD 128
QY 64 ELSAWKTPDSQTKKLMKACRMILQENELGRLSQGRIAGLEAEALQKYSSELKSS 123
DB 129 NVAANFTSSKLGKMLAKCRTLOEENEIGTMASEGKIHELGMKIAVLKSNQNELRNQ 198
QY 124 QDELNDFFIQLDEEVEGMSQSTILVLQQLKQKTRQQLAQVQOOQ-SQASAPSTSRRTTASEP 183
DB 189 FDVLYKHMGGVNDVRSNELVAILQEELEAKELEVLRLKEETLSQKGTPEAPAP 242
QY 184 EQSEATSKDCSRLTNGPNSGSSSRQ 208
DB 243 E-----TSDAG---NGQEADSDTLQ 260

RESULT 6

US-10-424-599-240261
; Sequence 240261, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684

; SEQ ID NO 240261
; LENGTH: 343
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_58983C.1.pep
US-10-424-599-240261

Query Match 16.4%; Score 264; DB 15; Length 343;
Best Local Similarity 32.7%; Pred. No. 6e-10;
Matches 54; Conservative 47; Mismatches 64; Indels 0; Gaps 0;

QY 4 ATKQDIOECTTQIYLYKQVQPSVAQLRSTWVDPAINLFLKMGKELETKDKLEQAQN 63
DB 146 AKRQEIAELKSAVRDLKQVLPSPMQTRRLDPAIHFEFTRLKLVKEKIKELQD 205
QY 64 ELSAWKTPDSQTKKLMKACRMILQENELGRLSQGRIAGLEAEALQKYSSELKSS 123
DB 206 NVAANFTSSKLGKMLAKCRTLOEENEIGTMASEGKIHELGMKIAVLKSNQNELRNQ 265
QY 124 QDELNDFFIQLDEEVEGMSQSTILVLQQLKQKTRQQLAQVQOOQ-SQ 168
DB 266 FEGLQKHMEGLTNDVRSNELVAILQEELEAKELEVLRLKEETLSQKGTPEAPAP 310

RESULT 7

US-10-425-115-203631
; Sequence 203631, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 203631
; LENGTH: 195
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_1172C.1.pep
US-10-425-115-203631

Query Match 16.3%; Score 262.5; DB 17; Length 195;
Best Local Similarity 31.8%; Pred. No. 3.8e-10;
Matches 64; Conservative 46; Mismatches 78; Indels 13; Gaps 3;

QY 8 QDIOECTTQIYLYKQVQPSVAQLRSTWVDPAINLFLKMGKELETKDKLEQAQNLSA 67
DB 3 QBITELKSAVRDLKQVLPSPMQTRRLDPAIHFEFTRLKLVKEKIKELQDNVAA 62
QY 68 WKFTPDSTQTKKLMKACRMILQENELGRLSQGRIAGLEAEALQKYSSELKSSQDEL 127
DB 63 VNFPTSSKLGKMLAKCRTLOEENEIGTMASEGKIHELGMKIAVLKSNQNELRNQ 122
QY 128 NDFIQLDEEVEGMSQSTILVLQQLKQKTRQQLAQVQOOQ-SQASAPSTSRRTTASEP 187
DB 123 YKHMGGVNDVRSNELVAILQEELEAKELEVLRLKEETLSQKGTPEAPAP 173
QY 188 ATSKDCSRLTNGPNSGSSSRQ 208
DB 174 -TSDAG---NGQEADSDTLQ 190

RESULT 8

US-10-425-114-48881
; Sequence 48881, Application US/10425114
; Publication No. US20040034888A1

GENERAL INFORMATION:
 APPLICANT: Liu, Jingdong
 APPLICANT: Zhou, Yihua
 APPLICANT: Kovalic, David K.
 APPLICANT: Screen, Steven E.
 APPLICANT: Tabaska, Jack E.
 APPLICANT: Cao, Yongwei
 TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 FILE REFERENCE: 38-21(53313)B
 CURRENT APPLICATION NUMBER: US/10/425,114
 CURRENT FILING DATE: 2003-04-28
 NUMBER OF SEQ ID NOS: 73128
 SEQ ID NO 48881
 LENGTH: 213
 TYPE: PRT
 ORGANISM: Oryza sativa nipponbare
 FEATURE:
 OTHER INFORMATION: Clone ID: LIB3474-004-D12_FLI.pap
 US-10-425-114-48891

Query Match 15.9%; Score 256; DB 15; Length 213;
 Best Local Similarity 31.8%; Pred. No. 1.2e-09;
 Matches 55; Conservative 47; Mismatches 71; Indels 0; Gaps 0;

QY 4 ATKQDIQECTTQIYQLKVOQPSVAQLRSTWDPAINLFFLMKGELEOTKDKLEQAO 63
 DB 16 AKREQETAEKSAVRDLTKQLRPSPMOTRLLDPAIHEEFTRLKNLVEKEKKIKELQD 75
 QY 64 ELSAWKFTPDSTQTKKLVAKGEMLIQENBELGRQLSGRIQAELAEALOKKYSEELKSS 123
 DB 76 NVAAVNFTPSSKHGKMLMAKCRITQEENEEIGAMASEGKHELGMKIAVLKTRNNELRQ 135
 QY 124 QDELNDFIQIDEEVEGMSQSTILVLOQLKETRQQLAQYQQOQSOASAPSTS 176
 DB 136 FNELYKHMDELNDVRSNEWVAILQDELETKDVELRLKEMLAQKEATDENK 188

RESULT 9
 US-10-437-963-196776
 Sequence 196776, Application US/10437963
 Publication No. US20040133343A1
 GENERAL INFORMATION:
 APPLICANT: La Rosa, Thomas J.
 APPLICANT: Kovalic, David K.
 APPLICANT: Zhou, Yihua
 APPLICANT: Cao, Yongwei
 APPLICANT: Boukharov, Andrey A.
 APPLICANT: Barbaruk, Brad
 APPLICANT: Li, Ping
 TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
 TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 FILE REFERENCE: 38-21(53221)B
 CURRENT APPLICATION NUMBER: US/10/437,963
 CURRENT FILING DATE: 2003-05-14
 NUMBER OF SEQ ID NOS: 204966
 SEQ ID NO 196776
 LENGTH: 336
 TYPE: PRT
 ORGANISM: Oryza sativa
 FEATURE:
 NAME/KEY: unsure
 LOCATION: (1)...(336)
 OTHER INFORMATION: unsure at all Xaa locations
 FEATURE:
 OTHER INFORMATION: Clone ID: PAT_MRT4530_92596C.1.pap
 US-10-437-963-196776

Query Match 15.9%; Score 256; DB 16; Length 336;
 Best Local Similarity 31.8%; Pred. No. 2e-09;
 Matches 55; Conservative 47; Mismatches 71; Indels 0; Gaps 0;

QY 4 ATKQDIQECTTQIYQLKVOQPSVAQLRSTWDPAINLFFLMKGELEOTKDKLEQAO 63
 DB 139 AKREQETAEKSAVRDLTKQLRPSPMOTRLLDPAIHEEFTRLKNLVEKEKKIKELQD 198
 QY 64 ELSAWKFTPDSTQTKKLVAKGEMLIQENBELGRQLSGRIQAELAEALOKKYSEELKSS 123
 DB 199 NVAAVNFTPSSKHGKMLMAKCRITQEENEEIGAMASEGKHELGMKIAVLKTRNNELRQ 258
 QY 124 QDELNDFIQIDEEVEGMSQSTILVLOQLKETRQQLAQYQQOQSOASAPSTS 176
 DB 259 FNELYKHMDELNDVRSNEWVAILQDELETKDVELRLKEMLAQKEATDENK 311

RESULT 10
 US-10-310-154-593
 Sequence 593, Application US/10310154
 Publication No. US20030233670A1
 GENERAL INFORMATION:
 APPLICANT: Edgerton, Michael D.
 APPLICANT: Chomet, Paul S.
 APPLICANT: Adams, Thomas H.
 APPLICANT: Ruff, Thomas G.
 APPLICANT: Agarwal, Ameeta K.
 APPLICANT: Ahrens, Jeffrey E.
 APPLICANT: Ball, James A.
 APPLICANT: Banu, G.
 APPLICANT: Bell, Brin
 APPLICANT: Boddupalli, Raghava
 APPLICANT: Deikman, Jill
 APPLICANT: Deng, Molian
 APPLICANT: Dong, Jinzhao
 APPLICANT: Duff, Stephen M.
 APPLICANT: Galligan, Meghan M.
 APPLICANT: Hinchey, Brenda S.
 APPLICANT: Huang, Shihshieh
 APPLICANT: Johnson, G. Richard
 APPLICANT: Jung, Vincent
 APPLICANT: Kretzmer, Keith A.
 APPLICANT: Laccetti, Lucille B.
 APPLICANT: Lai, Chao-Qiang
 APPLICANT: Lee, Gary
 APPLICANT: Lin, Jie-Yi
 APPLICANT: Liu, Jingdong
 APPLICANT: Lu, Bin
 APPLICANT: Lu, Michael M.
 APPLICANT: Lund, Adrian
 APPLICANT: Madson, Linda L.
 APPLICANT: Malloy, Kathleen A.
 APPLICANT: McKiel, Christine L.
 APPLICANT: Miller, Philip W.
 APPLICANT: Padmavathi, Manthikanti
 APPLICANT: Parnell, Laurence D.
 APPLICANT: Start, William G.
 APPLICANT: Tennesen, Dan
 APPLICANT: Viciya, K.R.
 APPLICANT: Wang, Haiyuo
 APPLICANT: Xin, Zhanguo
 APPLICANT: Xu, Nanfei
 APPLICANT: Yang, Chunzhi
 APPLICANT: Zeng, Xiaoping
 APPLICANT: Zhang, Qiang
 APPLICANT: Zhao, Yajuan
 APPLICANT: Zhou, Li
 TITLE OF INVENTION: Gene Sequences and Uses Thereof in Plants
 FILE REFERENCE: 38-15(52796)B
 CURRENT APPLICATION NUMBER: US/10/310,154
 CURRENT FILING DATE: 2002-12-04
 PRIOR APPLICATION NUMBER: 60/337,358
 PRIOR FILING DATE: 2001-12-04
 NUMBER OF SEQ ID NOS: 736
 SEQ ID NO 593
 LENGTH: 352
 TYPE: PRT


```
; ORGANISM: Oryza sativa
US-10-310-154-593

Query Match      15.9%; Score 256; DB 14; Length 352;
Best Local Similarity 31.8%; Pred. No. 2.2e-09;
Matches 55; Conservative 47; Mismatches 71; Indels 0; Gaps 0;

QY 4 ATKGDIOECTTQIOYKQVQPSVAQLRSTWVDPAINFLLKMGKELEBCTDKLEQAQN 63
Db 155 AKREQEIATLKSARVRLKQLRPPSNQTRLLDPAIHFEFRLKULVEKEKKIKELQD 214
QY 64 ELSAWKFTPDSTGKMLMAKCRMLIQENQELGRLSQRIQAQLAEALQKYSSELKSS 123
Db 215 NVAANFTPSKKGKMLMAKCRTLQENEEIGAVASEGKIHELGMKIAVLKTRNNELRNQ 274
QY 124 QDELNDFIQLDEVEGMSQSTIQLQOLKEFTQQLAQYQQOQSASAPSTSR 176
Db 275 FNELYKHMGLTNDVERSNEWMVAILODELTKDVELRLUKMLAQKEATDENK 327

RESULT 11
US-10-425-114-55340
; Sequence 55340, Application US/10425114
; Publication No. US20040034886A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 55340
; LENGTH: 190
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: JC-GMLE01810068D05_FLI.pep
US-10-425-114-55340

Query Match      15.4%; Score 247; DB 15; Length 190;
Best Local Similarity 35.3%; Pred. No. 4.2e-09;
Matches 49; Conservative 40; Mismatches 50; Indels 0; Gaps 0;

QY 30 QLRSTWDPAINLFLKMGKELEQTKKLEQAQNELSAWKFTPDSTGKMLMAKCRMLIQ 89
Db 8 QARLLDPAVHEEFTRLNKLVEEKKKVKELQDNIAVSTPSKMGKMLMAKCRTLQE 67
QY 90 ENQELGRLSQRIQAQLAEALQKYSSELKSSQDELNDFIQLDEVEGMSQSTIQLVQ 149
Db 58 ENEEIGNQASEGKHGKMLALQKSONSLRNQFEGQLKHMEGLTNDVERSNEWMVLQ 127
QY 150 QOLKETROQLAQYQQOQSQ 168
Db 128 EKLEEKDRIQRLKHEIQ 146

RESULT 12
US-10-739-930-9291
; Sequence 9291, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53377)B
; CURRENT APPLICATION NUMBER: US/10/739,930
; CURRENT FILING DATE: 2003-12-18
```

```
; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 9291
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: GLYMA-23APR03-C6342_2.p
US-10-739-930-9291

Query Match      14.8%; Score 238; DB 17; Length 299;
Best Local Similarity 36.8%; Pred. No. 3e-08;
Matches 50; Conservative 35; Mismatches 51; Indels 0; Gaps 0;

QY 4 ATKGDIOECTTQIOYKQVQPSVAQLRSTWVDPAINFLLKMGKELEQTKKLEQAQN 63
Db 146 AKREQEIATLKSARVRLKVLKPPSMQSRLLDPAVHEEFTRLNKLVEEKKKVKELQD 205
QY 64 ELSAWKFTPDSTGKMLMAKCRMLIQENQELGRLSQRIQAQLAEALQKYSSELKSS 123
Db 206 NIAVSTPSKMGKMLMAKCRTLQENEEIGNQASEGKHGKMLALQKSONSLRSQ 265
QY 124 QDELNDFIQLDEVE 139
Db 266 FEGLOKMEGLTNDVE 281

RESULT 13
US-10-424-599-240263
; Sequence 240263, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285884
; SEQ ID NO 240263
; LENGTH: 310
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(310)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_58985C.1.pap
US-10-424-599-240263

Query Match      14.8%; Score 238; DB 15; Length 310;
Best Local Similarity 34.6%; Pred. No. 3.1e-08;
Matches 53; Conservative 42; Mismatches 56; Indels 2; Gaps 2;

QY 4 ATKGDIOECTTQIOYKQVQPSV-AQLRSTWVDPAINFLLKMGKELEQTKKLEQA 61
Db 146 AKREQEIATLKSARVRLKVLKPPSNQAAXKEXLLLDASPHFEFTRLNKLVEEKKKVKEL 205
QY 62 QNELSAWKFTPDSTGKMLMAKCRMLIQENQELGRLSQRIQAQLAEALQKYSSELK 121
Db 206 QDNIAAVSTPSKMGKMLMAKCRTLQENEEIGNQASEGKHGKMLALQKSONSLRSQ 265
QY 122 SSOELNDFIQLDEVEGMSQSTIQLVQOLKE 154
Db 266 NQFEGLOKMEGLTNDVERSNEWMVLQEKLE 298

RESULT 14
US-10-424-599-240259
; Sequence 240259, Application US/10424599
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

QM protein - protein search, using sw model

Run on: November 14, 2004, 12:26:59 ; Search time 39 Seconds
(without alignments)
542.447 Million cell updates/sec

Title: US-10-030-389-2

Perfect score: 1608

Sequence: 1 MRLATKGDIQECTIQYL.....GSRHVQGLDSSVNVQGSVL 319

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66319000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued_Patents_AA.*
- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
 - 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
 - 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
 - 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
 - 5: /cgn2_6/ptodata/1/iaa/PGTUS_COMB.pep.*
 - 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	357	22.2	151	2	US-08-933-750C-1
2	357	22.2	151	3	US-09-234-613-1
3	166	10.3	525	4	US-09-107-532A-5095
4	156	9.7	897	1	US-08-095-737-4
5	156	9.7	897	1	US-08-480-145-4
6	156	9.7	897	2	US-08-477-389-4
7	155	9.6	57	4	US-08-621-976-5165
8	144	9.0	461	1	US-08-186-222-2
9	144	9.0	475	4	US-09-248-796A-14568
10	144	9.0	716	3	US-09-219-983A-20
11	141	8.8	1742	4	US-09-386-962C-4
12	141	8.8	1742	4	US-08-386-959-4
13	141	8.8	3248	5	PCT-US95-16216-1
14	141	8.8	3248	5	US-09-914-259-24
15	140	8.7	1031	4	US-08-328-254-6
16	139	8.6	2482	1	US-09-538-092-1154
17	139	8.6	3210	4	US-08-910-925-4
18	137.5	8.6	703	3	US-08-910-925-4
19	137	8.5	1024	4	US-09-270-787-44973
20	136.5	8.5	717	3	US-08-910-925-1
21	134.5	8.4	1155	4	US-09-710-279-1780
22	134	8.3	645	4	US-09-919-172-41
23	131.5	8.2	896	1	US-08-095-737-2
24	131.5	8.2	896	1	US-08-480-145-2
25	131.5	8.2	896	2	US-08-477-389-2
26	131.5	8.2	896	4	US-09-538-092-1113
27	131.5	8.2	1857	4	US-09-917-254-91

Sequence 1084, Ap
Sequence 919, Appl
Sequence 3, Appli
Sequence 4, Appli
Sequence 3, Appli
Sequence 2084, A
Sequence 14401, A
Sequence 2, Appli
Sequence 2, Appli
Sequence 16333, A
Sequence 1078, Ap
Sequence 520, Appl
Sequence 17, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-08-933-750C-1
Sequence 1, Application US/08933750C
Patent No. 5932442
GENERAL INFORMATION:
APPLICANT: Lal, Preeti
APPLICANT: Hillman, Jennifer L.
APPLICANT: Bandman, Olga
APPLICANT: Shah, Purvi
APPLICANT: Au-Young, Janice
APPLICANT: Yue, Henry
APPLICANT: Guegler, Karl J.
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
NUMBER OF SEQUENCES: 98
CORRESPONDENCE ADDRESS:
ADDRESSER: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/933,750C
FILING DATE: September 23, 1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0356 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 151 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: U937NOT01

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

Db 202 TSLAAEQATADKADLNKQ-KAEAEAEQARIREQORLAEQARQAAQAEKAEAREQAE 260
QY 167 SQASABSTGTTASEPVEQSEATSKDCSLRTNGPSNGSSRRORTSGSGFHREGNTTDDF 226
Db 261 AEAQATQASTAQSSESSATQSSMTSESSATQSSATESTT-----PSSSTEESTA 315
QY 227 PSSPGNGKSSNSRERTGRGSGYVNLQAGYSDPTGSENLSLTHQSN-----DT 279
Db 316 PES--SATESTTAPESSEATESTTVPSSATEESTVP---ESSSTEESTTAPPTTPT 370
QY 280 DSHD-----PQEEKAV-----SGKN-----RTVGRHVQNG 307
Db 371 DQSVDTGNGTSGSTPAPTPTPEQKXPVTPAPAPSGVNGAAVAEAYKIGTPIVWG 430
QY 308 LDSS 311
Db 431 KDPS 434

RESULT 4
US-08-095-737-4
; Sequence 4, Application US/08095737
; Patent No. 5487979
; GENERAL INFORMATION:
; APPLICANT: DiFiore, Pier P
; APPLICANT: Fazioli, Francesca
; TITLE OF INVENTION: A Substrate for the Epidermal Growth
; TITLE OF INVENTION: Factor Receptor Kinase
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: United States of America
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/095,737
; FILING DATE: 19930722
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned A
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH060.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 897 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-095-737-4

Query Match 9.7%; Score 156; DB 1; Length 897;
Best Local Similarity 23.7%; Pred. No. 0.00027;
Matches 79; Conservative 56; Mismatches 101; Indels 98; Gaps 16;

QY 3 LATKGODIQCETQIQYLK-QVQPSV-----AQLRSTWVDPAINLFLFKMKGELEQTKDK 57
Db 356 LKEKEDTVKQRTSEVDQLQDEVOREINLQKLQAKQKQVQ-----ELLGELDEQKQ 407
QY 58 LEQAQNELSAWKFTPDSTQTKKLMKACRMILQENQELGRQLSQGRITAEALQK--- 114
Db 408 LEEQLQK-----VRKKC-----AEEAQL-----ISSLKAEITSQESQI 440

QY 115 -KYSEELKSSODELNDP---LIQLDEEVEGMQSTILVLOQLKETROQLAQYQQQQSQAS 170
Db 441 SSYEELKKAEEELSRLOQETAQLSEESVSGKAELEFLQQLHLSQSEIISMMQRLM-- 498
QY 171 APSTSRITASEPVEQSEATSKDCSLRTNGPSNGSSRRORTSGSGFHREGNTTDDFPSP 230
Db 499 -----KDL-ETDNNQSNWSSSPQSVLVNG-----ATDYCSLS 529
QY 231 GNGKSSNSRERTGRGSGYVNLQAGYSDPTGSENLSLTHQSN-----DT 290
Db 530 TSSETANFNHAEQ-----NNLES-----EPHCESSV-RSSPEIAPSDVTDESEA 576
QY 291 VSGKGNRTVGR-----HVQNGLDSSVNVQSVL 319
Db 577 VTVAGNEKVTRFDDDKHKE--EDPFNVBSSL 608

RESULT 5
US-08-480-145-4
; Sequence 4, Application US/08480145
; Patent No. 5717067
; GENERAL INFORMATION:
; APPLICANT: DiFiore, Pier P
; APPLICANT: Fazioli, Francesca
; TITLE OF INVENTION: A Substrate for the Epidermal Growth
; TITLE OF INVENTION: Factor Receptor Kinase
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: United States of America
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,145
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/095,737
; FILING DATE: 22-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned A
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH060.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 897 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-480-145-4

Query Match 9.7%; Score 156; DB 1; Length 897;
Best Local Similarity 23.7%; Pred. No. 0.00027;
Matches 79; Conservative 56; Mismatches 101; Indels 98; Gaps 16;

QY 3 LATKGODIQCETQIQYLK-QVQPSV-----AQLRSTWVDPAINLFLFKMKGELEQTKDK 57
Db 356 LKEKEDTVKQRTSEVDQLQDEVOREINLQKLQAKQKQVQ-----ELLGELDEQKQ 407
QY 58 LEQAQNELSAWKFTPDSTQTKKLMKACRMILQENQELGRQLSQGRITAEALQK--- 114
Db 408 LEEQLQK-----VRKKC-----AEEAQL-----ISSLKAEITSQESQI 440

QY 115 -KYSBELKSSQDELND---IIQLDEVEGMQSTILVLOOQLKETROQLAQYQQQSQAS 170
 Db 441 SSBELKAREELRLQOETAQLAESVESGKAQLEPLQCHLQESQEISSNQMLEM-- 498
 QY 171 APSTRTTASBPVEOSEATSKDCSRLTNGPNSGSSSRQRTSGSFHREGNTTDDFPSP 230
 Db 499 -----KDL-ETDNNQSNWSSSPQSVLVNG-----ATDYCSLS 529
 QY 231 GNGKSSNSSBRTGRGGSGYVNLASAGYSDVPTGSENLSLTHQSDNTDSSHDPOBEKA 290
 Db 530 TSSSTANFNHAEQ-----NNLES-----EPHQESSV-RSSPEIAPSDVTDESEA 576
 QY 291 VSGKGNRTVGSR-----HVQNLGSSVNVQGSVL 319
 Db 577 VTVAGNEKVTFRFDDDKHSKE--EDPFNVESSSL 608

RESULT 6
 US-08-477-389-4
 ; Sequence 4, Application US/08477389
 ; Patent No. 5872219
 ; GENERAL INFORMATION:
 ; APPLICANT: Difiore, Pier P
 ; APPLICANT: Fazioli, Francesca
 ; TITLE OF INVENTION: A Substrate for the Epidermal Growth
 ; TITLE OF INVENTION: Factor Receptor Kinase
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Knobbe, Martens, Olson & Bear
 ; STREET: 620 Newport Center Drive, Sixteenth Floor
 ; CITY: Newport Beach
 ; STATE: California
 ; COUNTRY: United States of America
 ; ZIP: 92660
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/477,389
 ; FILING DATE: 07-JUN-1995
 ; CLASSIFICATION: 530
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/095,737
 ; FILING DATE: 22-JUL-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Israelsen, Ned A
 ; REGISTRATION NUMBER: 29,655
 ; REFERENCE/DOCKET NUMBER: NIH060.001A
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (619) 235-8550
 ; TELEFAX: (619) 235-0176
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 897 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-477-389-4

Query Match 9.7%; Score 156; DB 2; Length 897;
 Best Local Similarity 23.7%; Pred. No. 0.00027;
 Matches 79; Conservative 56; Mismatches 101; Indels 98; Gaps 16;
 QY 3 LATKGQDIQCTTQIQVLYX-QVQPSV-----AQLRSTWVDPAINLFLKMKGELEQTKDK 57
 Db 356 LKKEKEDVTQRTSEVDQLQDEQVESINLQKQAQQVQ-----ELLGELDEQRAQ 407
 QY 58 LEQAQNELSAWKFTPSQTGKLMKRCMLIQENQELGRQLSQGRIAQLAEALQK--- 114
 Db 408 LERLQLE-----VRKKC-----ABEAQL-----ISSLKARITSQESQI 440

QY 115 -KYSBELKSSQDELND---IIQLDEVEGMQSTILVLOOQLKETROQLAQYQQQSQAS 170
 Db 441 SSBELKAREELRLQOETAQLAESVESGKAQLEPLQCHLQESQEISSNQMLEM-- 498
 QY 171 APSTRTTASBPVEOSEATSKDCSRLTNGPNSGSSSRQRTSGSFHREGNTTDDFPSP 230
 Db 499 -----KDL-ETDNNQSNWSSSPQSVLVNG-----ATDYCSLS 529
 QY 231 GNGKSSNSSBRTGRGGSGYVNLASAGYSDVPTGSENLSLTHQSDNTDSSHDPOBEKA 290
 Db 530 TSSSTANFNHAEQ-----NNLES-----EPHQESSV-RSSPEIAPSDVTDESEA 576
 QY 291 VSGKGNRTVGSR-----HVQNLGSSVNVQGSVL 319
 Db 577 VTVAGNEKVTFRFDDDKHSKE--EDPFNVESSSL 608

RESULT 7
 US-09-621-976-5165
 ; Sequence 5165, Application US/09621976
 ; Patent No. 6639063
 ; GENERAL INFORMATION:
 ; APPLICANT: Dumas Milne Edwards, J.B.
 ; APPLICANT: Jobert, S.
 ; APPLICANT: Giordano, J.Y.
 ; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
 ; FILE REFERENCE: GENSET.054PR2
 ; CURRENT APPLICATION NUMBER: US/09/621,976
 ; NUMBER OF SEQ ID NOS: 19335
 ; SOFTWARE: Patent.pm
 ; SEQ ID NO 5165
 ; LENGTH: 57
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-09-621-976-5165

Query Match 9.6%; Score 155; DB 4; Length 57;
 Best Local Similarity 86.8%; Pred. No. 8e-06;
 Matches 33; Conservative 2; Mismatches 1; Indels 2; Gaps 1;
 QY 26 PSVAQ--LRSTWVDPAINLFLKMKGELEQTKDKLEOA 61
 Db 20 PRLAENKLRSWVDPAINLFLKMKGELEQTKDKLEOA 57

RESULT 8
 US-08-186-222-2
 ; Sequence 2, Application US/08186222
 ; Patent No. 5559007
 ; GENERAL INFORMATION:
 ; APPLICANT: Surli, Bruno
 ; APPLICANT: Schmitz, Albert
 ; TITLE OF INVENTION: Bacterial Vectors
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: CIBA-GEIGY Corporation
 ; STREET: 7 Skyline Drive
 ; CITY: Hawthorne
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10532
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/186,222
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/672,205

```
; FILING DATE: 19-MAR-1991
; APPLICATION NUMBER: GE 9006400.7
; FILING DATE: 22-MAR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Villamizar, JoAnn
; REGISTRATION NUMBER: 30,598
; REFERENCE/DOCKET NUMBER: 4-17994/A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914)785-7121
; TELEFAX: (914)347-5769
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 461 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-186-222-2

Query Match          9.0%; Score 144; DB 1; Length 461;
Best Local Similarity 22.0%; Pred. No. 0.001;
Matches 65; Conservative 53; Mismatches 124; Indels 54; Gaps 9;

QY 2 RLATGQDIQECTTQIQYLKQVQPSVAQLRSTWVDPAINLFFL-----KMKGE 50
DB 88 QIATINESIKERTKLE--AARSQVNSATNYDAVNSKSLTDVIQKVTATVSSA 145
QY 51 LEQTKDLQAQNELSAWFTPDSTQTKK----LMAKRMLIQENQELGROLSQGRIAQL 106
DB 146 NKMLFQEQKEQKELLS-----QKSTVKNYQNFVLSQSLDSQAQELTSQAELKVATL 200
QY 107 EABLALQKYSBELKSSQDELNDFIQLDEEVEGMCSTILVLQOQLKETROQLAQYQQQQ 166
DB 201 -----NYQATIATAQDKKQAL--LDKA-----AAEKAAQEAACKQAAYEAQQ 241
QY 167 SQASAPSTERTTASEPVEOSEATSKDCSLRLNPNPNSGSSSQRTSGSGFHREGNTTDDF 226
DB 242 KEAAQAQAATA-----ATAKAVEAATSSASASSAPQVSTST--TDNTTGNAS 288
QY 227 PPSGPNKNSNSSEERTGRGSGYVNSLSAGYSDVDSFTGSENSLTHOSNDTDS 282
DB 289 ASNSNSNSNSNSNSNSNSNSNSNSNSNSNSNSNSNSNSNSNSNSNSNSNSNSNSNS 344

RESULT 9
US-09-248-796A-14568
; Sequence 14568, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 14568
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-14568

Query Match          9.0%; Score 144; DB 4; Length 475;
Best Local Similarity 22.9%; Pred. No. 0.0011;
Matches 75; Conservative 53; Mismatches 132; Indels 68; Gaps 14;

QY 19 YLKQVQQPSVAQLRS-----TWVDPAINLFFLKMKGELFQTKDLKLEQAQNELSAW 68
DB 112 YTPSSQPSGERSLSQRAPSSNDLLADPAIS-----COLSQATSDIANVNSQIKSL 164

; FILING DATE: 19-MAR-1991
; APPLICATION NUMBER: GE 9006400.7
; FILING DATE: 22-MAR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Villamizar, JoAnn
; REGISTRATION NUMBER: 30,598
; REFERENCE/DOCKET NUMBER: 4-17994/A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914)785-7121
; TELEFAX: (914)347-5769
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 461 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-186-222-2

Query Match          9.0%; Score 144; DB 3; Length 716;
Best Local Similarity 18.6%; Pred. No. 0.0019;
Matches 78; Conservative 66; Mismatches 113; Indels 162; Gaps 13;

QY 5 TKGQDIQECTTQIQYLKQVQPSVAQLRSTWVDPAINLFFLKMKGELFQTKDLKLEQAQ 63
DB 339 SKNSLSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQ 385
QY 64 ELSAMKFTPDSTQTKKLMKRMKLIQENQEL--LGRQLSQGRIRIAQLEAELALQKYSLELKS 122
DB 386 ELQT-----QSDSQPRHLYQNLQNLKELSYQLAKLKELED-----QKKSQAEISK 436
QY 123 SQDELNDFII-----132
DB 437 GSNPSNLFIIQLPSEKPAFNGQPSIEPKLVQPGLSKPLPSGGGLIGKIPASTGLYL 496
QY 133 -----QLDEEVEGMCSTILVLQOQLKETROQLAQY---QQCQSQ-----168
DB 497 SPDFNDLSYRDQFLQQLQELAKKHQNLISLLQRRNDKKQKQKQKQKQKQKQKQKQK 556
QY 169 -----ASAPSTERTTASEPVE-----QSEATSKDCSLRLNPNPNSGSSR-----207
DB 557 KQSSSSAGSSCTKLQDQISTGAGSQSQQLQAGSTGLQTSLSLQGTSSASQALORLKE 616
QY 208 -----QRTSGSGFHREGNTTDDFPSSPGNKGNSNSSEERTGRGSGYVNSLS 256
DB 617 QELQRIQTENDQKTSSSSSSHNSQ-----NSQSSSSQSSQSSQSSQSSQSSQSSQSS 668
QY 257 AGYESVDSFTGSENSLTHOSNDTDSHDPOEKAQVSGKGNRTVGSRHVQNGLOSSVNVQ 315

RESULT 10
US-09-219-983A-20
; Sequence 20, Application US/09219983A
; Patent No. 6380159
; GENERAL INFORMATION:
; APPLICANT: Woliner, Mariana
; APPLICANT: Lung, Oliver
; APPLICANT: Tran, Khanh-Uyen
; TITLE OF INVENTION: GENES FOR MALE ACCESSORY GLAND PROTEINS IN DROSOPHILA
; FILE REFERENCE: 19603/1791
; CURRENT APPLICATION NUMBER: US/09/219,983A
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/071,315
; PRIOR FILING DATE: 1997-12-23
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-219-983A-20

Query Match          9.0%; Score 144; DB 3; Length 716;
Best Local Similarity 18.6%; Pred. No. 0.0019;
Matches 78; Conservative 66; Mismatches 113; Indels 162; Gaps 13;

QY 5 TKGQDIQECTTQIQYLKQVQPSVAQLRSTWVDPAINLFFLKMKGELFQTKDLKLEQAQ 63
DB 339 SKNSLSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQ 385
QY 64 ELSAMKFTPDSTQTKKLMKRMKLIQENQEL--LGRQLSQGRIRIAQLEAELALQKYSLELKS 122
DB 386 ELQT-----QSDSQPRHLYQNLQNLKELSYQLAKLKELED-----QKKSQAEISK 436
QY 123 SQDELNDFII-----132
DB 437 GSNPSNLFIIQLPSEKPAFNGQPSIEPKLVQPGLSKPLPSGGGLIGKIPASTGLYL 496
QY 133 -----QLDEEVEGMCSTILVLQOQLKETROQLAQY---QQCQSQ-----168
DB 497 SPDFNDLSYRDQFLQQLQELAKKHQNLISLLQRRNDKKQKQKQKQKQKQKQKQKQK 556
QY 169 -----ASAPSTERTTASEPVE-----QSEATSKDCSLRLNPNPNSGSSR-----207
DB 557 KQSSSSAGSSCTKLQDQISTGAGSQSQQLQAGSTGLQTSLSLQGTSSASQALORLKE 616
QY 208 -----QRTSGSGFHREGNTTDDFPSSPGNKGNSNSSEERTGRGSGYVNSLS 256
DB 617 QELQRIQTENDQKTSSSSSSHNSQ-----NSQSSSSQSSQSSQSSQSSQSSQSSQSS 668
QY 257 AGYESVDSFTGSENSLTHOSNDTDSHDPOEKAQVSGKGNRTVGSRHVQNGLOSSVNVQ 315
```

Db 669 -----LDQSSKTSRSP-----SRRLNRRHRRSORETHLO 703

```

RESULT 11
US-09-386-962C-4
; Sequence 4, Application US/09386962C
; Patent No. 6635473
; GENERAL INFORMATION:
; APPLICANT: FOSTER, Timothy
; TITLE OF INVENTION: POLYPEPTIDES AND POLYNUCLEOTIDES FROM COAGULASE-NEGATIVE STAPHYLOCOCCUS EPIDERMIDIS
; FILE REFERENCE: P6635473/BAS
; CURRENT APPLICATION NUMBER: US/09/386,962C
; CURRENT FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: 60/038,443
; PRIOR FILING DATE: 1998-08-31
; PRIOR APPLICATION NUMBER: 60/117,119
; PRIOR FILING DATE: 1999-01-25
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 1742
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-386-962C-4

```

Query Match	8.8%;	Score 141;	DB 4;	Length 1742;	
Best Local Similarity	24.3%;	Pred. No. 0.011;			
Matches	74;	Conservative 48;	Mismatches 124;	Indels 58;	Gaps 13;
Qy	41	NLFF-----LKMGELEQTKLEQAQNELSAWKTPTDSQTGKKLMAKRMUJQENQE	93		
Db	1	NIYFNWRYSMKKRQGPINRVDFLSKNVKYRSIKET-----VGTASILVGATLM	51		
Qy	94	LGROLSGRIQAQLEALOKKYSEELKSSQDELNDLFIQLDEVEGQWQSTILVLOOQLK	153		
Db	52	FGAADNAKAAE---DNQLSASKEBKQGRDNENSKLNQVLDL-NGSHS-----EK	100		
Qy	154	ETRQLQAOYQQQSOQASAPT---SRITAGPEVQSEATSKDCSRLTNGSPNGSSSRQRT	210		
Db	101	TTNVNNA---TEVKVZEAPTTSDVSKPKANEAVVTNESTK---PKTTEATVNEESAET	154		
Qy	211	SGSGFPHREGNTEDDDPSSPGNGKNSNSSEERTGRCGSGYVNLQLS---AGYSEVDSQPTG	267		
Db	155	PKSTSTTQQ-DSTENKNPSLNDLNSSTSTTSKESKTDEHSTKQAQMSTKNKSLNDTNSPTQ	213		
Qy	268	SENSLTHQSNDDTSSHD-----PQEBKAVSGKCN--RTVGSRHVQNGL-----D	309		
Db	214	SEKTSSCANNDSTNQSAPSKQLDSKPSQKVYTKFNDEPTQDVEHTTTKLKTPSVSTD	273		
Qy	310	SSVN 313			
Db	274	SSVN 277			

```

RESULT 12
US-09-386-959-4
, Sequence 4, Application US/09386959
, Patent No. 6703025
, GENERAL INFORMATION:
, APPLICANT: PATTI, Joseph M.
, APPLICANT: FOSTER, Timothy J.
, APPLICANT: HOOK, Magnus
, TITLE OF INVENTION: MULTICOMPONENT VACCINES
, FILE REFERENCE: P0633303US/BAS
, CURRENT APPLICATION NUMBER: US/09/386,959
, CURRENT FILING DATE: 1999-08-31
, EARLIER APPLICATION NUMBER: 607098,439
, EARLIER FILING DATE: 1999-08-31
, NUMBER OF SEQ ID NOS: 65
, SOFTWARE: Patent In Ver. 2.0
, SEQ ID NO 4
, LENGTH: 1742

```

```

; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-386-959-4

```

```

Query Match      8.8%; Score 141; DB 4; Length 1742;
Best Local Similarity 24.3%; Fred.No. 0.011;
Matches 74; Conservative 46; Mismatches 124; Indels 58; Gaps 13;

Qy 41 NLFF-----LXWKGLEOTKQLDAQNELSANWETPDQSGTKMLAKRMLIQEQE 93
Db 1 NIFPNWYSWKRRQGGPKINRVDFLSNVKNKYSIRKPT-----VGTASILVGATLM 51
Qy 94 LGRQLSGRIQALEAEIALQKYSEELKSCQDELNDFIIQLDEEVEGMQSTILVLQOOLK 153
Db 52 FGAADNEAKAE---DNQLESASEKEQKGRDNENSKLNQVDLD-NGSHS-----EK 100
Qy 154 ETQQLQIAYQOQOSQASAPST---SRTTASEPVEQSEATSKDCSRJLNGPNSGSSRRQT 210
Db 101 TTNVNA--TEVKYVEAPTTSDVSKPKANEAVVTNESTK---PKTTEAPTVNEESIAET 154
Qy 211 SGSGPFRHGNTTDEDDPSPGNGCNKSNSEERTGRGSGYVNLQSL---AGYRSVDSPTG 267
Db 155 PKTSTTQQ-DSTEKNNPFLKDNLSSTTSKESKTDHSTKQAQMNTKSNLNDTNDSPQT 213
Qy 268 SENSLTHQSNDDTDSHD-----PQEKAVSGKGN--RTVGSHSVQNGL-----D 309
Db 214 SEKTSQANDSTDNQSAKQLDSEKPEQKVYKTFENDEPTQDVZHTTKLXTPSVSTD 273
Qy 310 SSVN 313
Db 274 SSVN 277

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999

```



```
;
;
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HUMAN
; US-08-353-700-1

Query Match      8.8%; Score 141; DB 1; Length 3248;
Best Local Similarity 22.4%; Pred. No. 0.025;
Matches 73; Conservative 61; Mismatches 118; Indels 74; Gaps 15;

QY 12 ECTTQIYLVKQVQPSVAQLRSTWDPAINLFLKMKGELEQTKDLEQAQNELSAWKFT 71
DB 2857 KATTOI-----LEELKTKYD--NLKY-----VNQLKENERAQGM-----2891

QY 72 PDSOTGKLMKAKRMILQENQELGRQLSQRIQAQLEA--LALOKKYSE---ELKSSQDE 126
DB 2892 -----KLLIKSCQLBEKEILQKLSQLQAQEKQKTGYMDTKVDLTTTEIKELKET 2945

QY 127 LNDFFIQLDEVEGQSTILVLOOQLKETROQL-----AQYQQQSQASAPSTSRIT 178
DB 2946 LEEKTKEADEYLD-KYCSLLISHEKLEKAKEMLETVQVHLCSQQSKDSRGSPLLGVPVP 3004

QY 179 ASEPVQSEATSKDCSRLTNGPSNGSSSRORTSGSGFHREGNTTDDFFSPGNGKSN 238
DB 3005 GPSPI--PSVTEK--RLSSGQNKASGKRQRSSGIWENGRT-----PATPESFKSK 3054

QY 239 SS-----BERTGRGGSGYVNLQLSAGYVSDPTGSENLSLTHOSNDTSSHDPQ--286
DB 3055 KAVMSGIHPAEDTEGTEFEPEGLPEVVKKGK--ADIPTGKTSPIYLRTTMTATRTSPRLA 3112

QY 287 -EEKAVS-----GKGNRTVGSRHVQNG 307
DB 3113 AQKLAISPLSLGKLENLAESSKPTAGG 3138

RESULT 14
PCT-US95-16216-1
; Sequence 1, Application PC/TUS9516216
; GENERAL INFORMATION:
; APPLICANT: Yen, Timothy J.
; APPLICANT: Rattner, Jerome B.
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16216
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,700
; FILING DATE: 09-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid

;
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; PCT-US95-16216-1

Query Match      8.8%; Score 141; DB 5; Length 3248;
Best Local Similarity 22.4%; Pred. No. 0.025;
Matches 73; Conservative 61; Mismatches 118; Indels 74; Gaps 15;

QY 12 ECTTQIYLVKQVQPSVAQLRSTWDPAINLFLKMKGELEQTKDLEQAQNELSAWKFT 71
DB 2857 KATTOI-----LEELKTKYD--NLKY-----VNQLKENERAQGM-----2891

QY 72 PDSOTGKLMKAKRMILQENQELGRQLSQRIQAQLEA--LALOKKYSE---ELKSSQDE 126
DB 2892 -----KLLIKSCQLBEKEILQKLSQLQAQEKQKTGYMDTKVDLTTTEIKELKET 2945

QY 127 LNDFFIQLDEVEGQSTILVLOOQLKETROQL-----AQYQQQSQASAPSTSRIT 178
DB 2946 LEEKTKEADEYLD-KYCSLLISHEKLEKAKEMLETVQVHLCSQQSKDSRGSPLLGVPVP 3004

QY 179 ASEPVQSEATSKDCSRLTNGPSNGSSSRORTSGSGFHREGNTTDDFFSPGNGKSN 238
DB 3005 GPSPI--PSVTEK--RLSSGQNKASGKRQRSSGIWENGRT-----PATPESFKSK 3054

QY 239 SS-----BERTGRGGSGYVNLQLSAGYVSDPTGSENLSLTHOSNDTSSHDPQ--286
DB 3055 KAVMSGIHPAEDTEGTEFEPEGLPEVVKKGK--ADIPTGKTSPIYLRTTMTATRTSPRLA 3112

QY 287 -EEKAVS-----GKGNRTVGSRHVQNG 307
DB 3113 AQKLAISPLSLGKLENLAESSKPTAGG 3138

RESULT 15
US-09-914-259-24
; Sequence 24, Application US/09914259
; Patent No. 5495336
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914,259
; CURRENT FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 1031
; TYPE: PRT
; ORGANISM: Strongylocentrotus purpuratus
; US-09-914-259-24

Query Match      8.7%; Score 140; DB 4; Length 1031;
Best Local Similarity 21.3%; Pred. No. 0.0064;
Matches 70; Conservative 50; Mismatches 125; Indels 84; Gaps 11;

QY 2 RLATKGQDIOECTTQIYLVKQVQPSVAQLRSTWDPAINLFLKMK-----GEL 51
DB 710 QLANLRTNEKEHQWELKQVQ-----RMTLQHEKQLDYKLEKIEAEKAAKREL 763

QY 52 EOTKDKLEQAQNELSAWKFTPDSOTGKIL--MAKCRMLI-----QENQE 93
DB 764 SQQFDRREQAKQDLKGLBET---VAKELQTLNLRLKLFVSDLNVRKKALEGGRDDDS 819

QY 94 LGRQLSQRIQAQLEAELALQKYSBELKSSQDELNDFFIQLDEVEGQSTILVLOOQLK 153
DB 820 GGSQAQKQKISFLNNLEQLTKVHQLVDNADRLCELFLERRLRATSERVALEMSLK 879

QY 154 ETRQ-----QLAQYQQQSQASAPSTSRITTASEPVQSEATSKDCSRLTNGPSNGSSROR 209
```

```

Db      ||:::      :|||:      :      :|||:      :|||:
880 ETKGAMRDRKRYQOEVDRI-----REAVROKFAKRGSSAQIA 918

QY      :      :      :      :      :      :      :
210 TSGSGFHREGNTTDDFPSSPCNGKNSSEERTGRGSGYVNLGAGYESVDSPT-GS 268

Db      :      :      :      :      :      :      :
919 KAIRAGHP-----PSPFGS-----TGIRGGYSGIRGGSPVIRPPSHGS 959

QY      :      :      :      :      :      :      :
269 ENSLTHQSDTSSHDPOEEKAVSGKGNR 297

Db      :      :      :      :      :      :      :
960 PEPISH-NNSPEKSLNPDAENMEKKANK 987

```

Search completed: November 14, 2004, 12:37:14
 Job time : 41 secs